Moore's Moral Facts and the Gap in the Retributive Theory

***Criminal Law and Philosophy*, forthcoming. The final publication is available at www.springerlink.com : http://www.springerlink.com/content/d181863747r88302/**

I

The purely retributive moral justification of punishment –the view that offenders should be punished because they deserve it, and that this moral desert is both a necessary and a sufficient reason to punish- has a gap at its centre which makes it vulnerable to the criticisms of rival theories. It fails to explain how, in the case of convicted offenders, afflicting a fellow human being, causing suffering in the world that is that person's conscious life, becomes a good thing to do instead of a bad one. Consequentialist theories can appeal to the principle that we may do a bad thing (cause suffering and loss of amenity and opportunity to a convicted criminal) in order to achieve good things that outweigh it (diminish suffering for others, by deterring future crime or by incapacitating this offender, etc.). There are problems with these theories, but they do not include the problem that they represent causing suffering to a human being as a good thing to do. Rather, they appeal to the notion of the lesser of two evils. Retributive theories that eschew justification on grounds of reduced human suffering do face that problem.

Defenders of the pure retributive theory have a range of options for filling, or denying, the gap. They may look for ways of denying the distinction I have just drawn between the implications of consequentialist and pure retributive theories. Or if they accept

that the pure retributivist does indeed have a special problem, in seeming to violate a *prima facie* moral rule against the infliction of uncompensated suffering, they may try to solve that problem in several different ways. They will need to show that, within the sphere of criminal justice, the punishable person is excluded from the protection of this moral rule.[1] They must therefore either (i) make a claim about the person prior to his entry into the criminal justice process, such that in virtue of his offending act he has put himself outside the scope of the protecting rule; (ii) make a claim about the sphere of criminal justice itself, such that it is a consequence of that claim that the protecting rule gives way to some superior (but non-consequentialist) good; or (iii) make a claim that knocks out the very intuition that motivates the protecting rule itself, that is with a more powerful intuition specially invoked by the situations to which criminal justice is addressed. The forms these claims typically take may be called the *theory of the offender's forfeited right*, the *intrinsic good of justice theory*, and the *moral facts theory*.

In this paper, there will only be space to address (iii), in the version powerfully argued by Michael S. Moore in the essays collected in *Placing Blame*. Moore argues that certain emotions that sane and decent people feel in response to serious criminal acts are caused by, and therefore provide evidence of, moral facts.

> The moral fact of the matter often causes our moral beliefs though the intermediate causing of our emotional responses. Our emotions in such case become good evidence of the underlying moral landscape…. Far from

[1] By '*prima facie* moral rule' I do not mean a positive rule generated within a worked-out moral system. I mean rather 'the kind of general imperative we are likely to reach for when we pre-theoretically formulate our moral ideas'. That there is a pre-theoretical moral inhibition against inflicting suffering will not, I imagine, be denied.

> hampering our insights into the truth, our emotions are often our best route to discovering the truth.[2]

These moral facts, which point to the truth of retributivism, form, according to Moore, essential elements of a valid moral theory underpinning the justification of punishment. My aim in this paper will be to call Moore's claim into question.

II

Moore's strategy for vindicating retributivism may be summarised as follows. First, he maintains that criminals can be held morally responsible for their actions. Let us accept this. Secondly, he equates 'moral responsibility' with 'desert', explicitly at one point,[3] and implicitly throughout. This implication is important, since it is itself an attempt to bridge the gap: 'moral responsibility' describes an aspect of an agent's relation to an action that she does, while 'desert' describes (instead, or also) an aspect of an agent's relation to something that she may receive or may be done to her. Speaking informally, 'moral responsibility' looks backward to the act, and 'desert' looks both backward to the act and forward to the punishment.

At another point Moore equates 'desert' with 'culpable wrongdoing' (in respect of blameworthy actions).[4] But the introduction of 'culpable' does not help bridge the gap.

[2] Moore 1997, 183.

[3] Moore 1997, 91.

[4] Moore 1997, 168.

If it simply means ‘in a way that makes the agent morally responsible’, it leaves the gap unbridged; if it is also intended to include the idea ‘eligible to be punished’, then it bridges the gap with its ambiguity. In short, Moore wants us to accept that from the fact that an agent has been responsible for a certain action, it follows that it is right that others perform certain other actions. But this plainly does not follow unless some intermediate premises are established. The argument required would run as follows.

M1 X is morally responsible for a serious crime [which afflicts others].

M2 One who is morally responsible for a serious crime [which afflicts others] deserves affliction by the state, in the form of punishment.

M3 The state ought to afflict, in the form of punishment, those who deserve affliction on the grounds mentioned in M2.

Therefore,

M4 The state ought to afflict X, by reason of M1 – M3.

The phrase in square brackets allows the argument to avail itself of the claim that the person-afflicting nature of crime forms at least part of the justification for the person-afflicting nature of punishment. Most of Moore’s discussion implies that he believes this, and it is a common view among popular justifications of retributivism. I will

assume from now on that it is in play, though it would be logically possible for a pure retributivist to do without it.[5]

M2 and M3, taken together, attempt to bridge the gap. We will deny M2 if we are unconvinced that the notion of X's 'deserving', whether or not explained with reference to his earlier affliction of Y or Z, provides an intelligible basis for justifying an affliction by the state of X. We will deny M3 if we deny M2, or if we believe that, even if criminals do deserve to be afflicted, the state should not act according to this desert, but in some other way.[6]

The evidence for M2 is provided, in Moore's view, by certain moral facts. These are disclosed to us by the judgments with which well-adjusted people respond to acts of crime and punishment, and which according to Moore serve as a reliable heuristic to the moral properties of those acts. Such judgements are not mere 'untutored intuition'. They are susceptible of rational correction according to such criteria as consistency and coherence, and are also valid only when conditioned by virtue: "the touchstone of our epistemically reliable emotion (as opposed to a kind of moral hallucination) is the virtue of feeling such an emotion."[7] But subject to these quality-control correctives, we are asked to accept that such judgements provide strong evidence of moral facts, such as the fact that one who is morally responsible for a serious crime deserves punishment, and that this desert is both a sufficient and a necessary reason to punish.

[5] For example, one could believe that all lawbreakers deserve to be afflicted even if no-one is harmed by lawbreaking: the intrinsic worth of the law would 'require' that those who break it to be caused to suffer.

[6] The restriction to serious crimes in the argument is important. Moore himself, though willing in principle to regard all immoral conduct "no matter how slight" as eligible for punishment, accepts that the retributive good of punishing "minor moral wrongs" may be overridden by other goods, such as "the liberal goods of pluralism, tolerance and autonomy" (187).

[7] Moore 1997, 229.

Though Moore places particular weight on the compelling force of virtuous retributive emotions aroused by violent crimes, he does not rely entirely on the significance of these powerful emotions: he also invokes analogies with our understanding of sanctions and entitlements in somewhat less inflammatory legal contexts, such as tort and property rights. In these contexts we are often guided, according to Moore, by a belief that *desert* is sufficient to justify the maintenance of a right or the imposition of a sanction. For example, "we think that the person who works hard to produce a novel deserves the right to determine when and under what conditions the novel will be copied for others to read".[8] To assess the independent force of these analogies, which play at most a supporting role in Moore's development of his theory, would require a separate paper; but they provide an opportunity to make further important distinctions among conceptions of desert. If we reflect on Moore's authorship example, we can see that the notion of desert or deserving as applied to a citizen under the law has at least three senses: (i) desert as a relation internal to an existing institution, (ii) desert as a principle that motivates and justifies an institution, and (iii) what I will call intrinsic personal moral desert, independent of any institutional form –roughly, being a good person or one who has done what is morally good in the relevant context. As an example of (i), it is true that, since we have an institution of intellectual property within which certain rights are granted on the basis of authorship, an author may be said within the discourse of that institution to 'deserve' to enjoy the benefit of those rights, while someone who has infringed them may 'deserve' to be obliged to pay damages or destroy the infringing publications. In this case, what we call the desert of the author is just another name for

[8] Moore 1997, 113.

her positive right. We do not make any special moral judgement on her case, since the rule we have already collectively established is sufficient to determine her rights. As an example of (ii), we may believe it to be desirable that authors should be acknowledged and rewarded for the labour of authorship, and so establish an institution of intellectual property rights. In this case, what we call the desert of authors need only express our collective belief that our law will serve us well if it allows those of us who write books and can find publishers to affirm their authorship publicly and to earn income through the sale of their work. This belief would be sufficient to motivate us in establishing the institution, and to justify it morally.[9]

Both these senses of 'desert' contain a moral element: (i), though it involves no case-specific moral assessment, applies the general moral justification of (ii) to particular cases. Thus the claim that the intellectual property of an author involves 'desert' is true in the sense that (ii) intellectual property law has a moral justification, and (i) according to that law, she owns the intellectual property in her work: she may therefore be said to 'deserve' copyright protection, and the moral overtones of 'deserve' are not inappropriate because the institution that confers her right is a morally justified one. But this moral language need not imply any attribution of (iii), intrinsic personal moral desert in the relevant context, to the author, let alone imply that it is *because and only because* of that intrinsic personal moral desert, or the copyright-infringer's lack of it, that she should enjoy the fruits of her intellectual property. However, for the purposes of Moore's pure retributivist argument in criminal law, it is this third conception, intrinsic personal moral desert, that must be

[9] It is not, we should note, a universal belief forced on us by first principles of reason. It is culturally determined: some past societies have had no such conception of authorship; while other grounds for property entitlements, such as inheritance, are morally quite controversial even in modern Western culture.

invoked. It is uncontroversial that the institution of criminal law has some sort of moral justification, and that an offender may be said relative to this institution to 'deserve' punishment. Moore's controversial claim is that an intrinsic, extra-institutional kind of desert provides a reason for us to find the moral justification of punishment in giving offenders what they deserve.

III

In the most compelling passages of his intuition-based argument, Moore uses powerful examples to remind us that we very much want to afflict certain offenders, and to convince us that this desire reveals a moral fact. One, originally cited by Jean Hampton, is taken from Dostoevsky's *The Brothers Karamazov*. Ivan Karamazov reports the story: a nobleman imprisons an eight-year-old boy who has accidentally hurt one of his dogs, and then the following morning unleashes the entire pack to tear the boy apart before the eyes of his mother. Should we shoot the nobleman "for the satisfaction of our moral feelings"? Ivan mockingly asks. The normally gentle Alyosha responds with the words, "Shoot him!"[10] (Moore remarks that "Kant would approve" this answer.) Dostoevsky's own implication is not actually as clearly retributive as Moore's use of the example might suggest, but the episode will undoubtedly prompt most readers to share Alyosha's response. What are we to make of this widespread emotional response, and the claim that it provides evidence of a moral fact?

[10] Moore 1997, 99 (n. 30); Cf. Murphy and Hampton 1998, 111. The episode occurs in Book 5, chapter 4 of *The Brothers Karamazov*.

Moore's attempt to win support for pure retributivism from such examples appeals partly to their sheer emotional –and he would claim, cognitive- power: as he says, they "work our intuition pumps… vigorously".[11] We just know what we should do, some pure retributivists might say, thanks to these intuitions: we have reached the bedrock of moral necessity and further reasons cannot and need not be given. But Moore also recognises the advantage of situating these responses within a theory that explains why we should pay attention to them.

As Moore is aware, the claim that emotions can provide evidence of moral reality is open to two related objections: first, that the *modus operandi* of such revelation needs a lot of explanation, since we do not usually look to our emotions for reliable information; and second, that certain compelling emotions (for example, the righteous indignation of a lynch mob) seem to be 'moral hallucinations', offering apparently irresistible justifications for actions that in a calmer moment we would judge to be morally wrong. Moral guidance from our emotions looks both puzzling and risky. Nevertheless, that there is *some* relationship between emotional response and moral judgement is suggested by the frequency with which both are energised by the same phenomena. Moore sets himself the very difficult task of showing that the relationship is such that emotions can provide evidence of moral truths, which in turn justify certain actions. I will argue that other theories of the relationship between emotions and morality are at least as plausible, so that in the last resort Moore has to rely on persuading the reader that the force of the Alyosha intuition is an irresistible tie-

[11] Moore 1997, 185.

breaker in favour of his view.[12] I will end by arguing that the intuition itself, which I share, does not justify a pure retributivist conception of punishment.

To begin with what I take to be common ground: we know that observed actions and events can give rise to emotional responses, and that such emotional responses may develop in a way that corresponds to moral judgements. For example: our horror at the injury sustained by a car crash victim may be followed by, or blended with, moral indignation if we believe that the recklessness of another driver has caused the crash. This moral indignation feels quite different from the mere feeling of horror. It also feels quite different from a dispassionate judgement that the reckless driver is criminally responsible. Neither the feeling of horror alone, nor the dispassionate judgement as to criminal responsibility alone, can reasonably be thought to tell us what to do with the reckless driver: both are compatible with a consequentialist view of punishment. It is this distinctive feeling of moral indignation that (if anything) impels us to a retributive intuition.

But if this emotion of moral indignation was itself just caused by the combination of reactive horror and judgement of culpability, it could add nothing to our understanding. The moral emotion, then, to serve Moore's purpose, must show us something new. What can this be? The moral emotion has –directly or indirectly- been caused, at least in part, by what happened; and this may tempt us to believe that it evidences moral features of the event itself. But since this is just what the lynch mob would claim, we need to look very carefully at the cognitive status of the emotion. I will set aside any appeal to the suspicion that the decisive cause of the

[12] See Matravers 2000, 81-87 for a related discussion; see also Duff 1996, 28-31.

retributive moral emotion lies in some discreditable region of the human psyche. Let us accept for the sake of argument that no unworthy instincts are in play. We should ask: what is the moral emotion, with its specific retributive content, supposed to be? Is it like a perception? Is it like a belief? Is it like a desire? If we cannot come up with a plausible answer to these questions, the claim that emotions can evidence moral reality remains obscure.

We should note first that neither the specific emotion nor the moral judgement can be *immediately* caused by the event. Rather, each is a reactive commentary upon it. We can see this by comparing them with the paradigm case of a mental state caused by, and therefore evidencing, a fact: a perception. We are not *compelled* to experience a given emotional response or moral intuition simply by brute facts, in the way that we *are* compelled to perceive a crashed car by the brute fact that the crashed car is in our visual field, with the consequence that the perception (in normal conditions) can be taken as evidence of the material fact. In the case of the visual experience, the belief that we are seeing a crashed car follows involuntarily upon the perception. Borrowing Searle's terminology, we can say that the condition of satisfaction of the visual experience is *that there is a crashed car there*. (Even if we were, as it happens, having a hallucination, we would still know that the condition *that there is a crashed car there* was the one that would need to be fulfilled for our supposed perception not to be a hallucination.)[13] But the moral judgement on the reckless driver, however quickly arrived at, is a complex act of reflection. Unless we already have a conception of moral responsibility, a set of moral values, and cultural practices of reproof and

[13] cf. Searle 1983, 38-41.

punishment, we will not be able to interpret what we see in a way that gives our emotion the distinctive character we can call retributive indignation.

As for the emotion itself, supposed by Moore to evidence moral reality, to mediate between the moral fact and the moral judgement, what are *its* conditions of satisfaction? On a view now widely taken to be discredited, there are no conditions of satisfaction for an emotion, because an emotion lacks intentional content, being in itself merely a blind and irrational impulse, triggered by events but incapable of telling us anything about them.[14] A pure retributivist could hold this view, but all that he could then infer from Alyosha's emotion, and ours, is that people generally get very upset as a result of actions like those of the nobleman. We knew that already, and this evidence of a reaction provides for no distinction between Alyosha's response and that of a lynch mob. For the Alyosha response to do some work for Moore, he needs it to have some illuminating cognitive content, in which case it must have conditions that would make it true or false, or in some other way satisfied or not satisfied. We could construe the emotion primarily as a desire, the condition of satisfaction being *that I, or someone, shoot the nobleman*. But the only reality for which that could provide evidence is the reality of our desire. If we construe the emotion as a kind of passionate belief or assertion, *that the nobleman's act is morally wrong*, it appears to be no more than an epiphenomenon of the moral judgement, since the condition that, if met, will make the moral judgement correct will also automatically 'satisfy' the emotion. What is really needed for the emotion to do any work for Moore is for it to *combine or link* these two intentional contents,

[14] See Kahan and Nussbaum 1996 for arguments against this 'mechanistic' view.

demonstrating (somehow) that the belief that the nobleman's act is morally wrong necessitates the desire and intention that someone shoot him.

I cannot see how the emotion can be characterised in terms of conditions of satisfaction so as to give this conveniently gap-bridging result. Undoubtedly there is a train of thought, with a significant emotional content, which runs something like this: *the nobleman's act (which greatly distresses me) is such that it would be proper (and I ardently desire) that somebody shoot him*. I think this is quite an accurate description of Alyosha's train of thought in so far as I empathise with it. But it is a puzzling claim that this train of thought could be caused by, and so provide evidence of, moral reality. The bridge between the nobleman's act and the prospective act of retribution, captured in the phrase *is such that it would be proper that* looks like the interpolation of a logical connection by a reasoning consciousness, or alternatively a gloss placed by that consciousness upon a retaliatory instinct, rather than something that could, so to speak, exist in nature and cause a corresponding thought to exist. If such a thought were a 'moral hallucination', would the person undergoing it –like the person with a hallucination of seeing a car- still know what condition would need to be fulfilled for it not to be a hallucination? For the 'perception hallucinator', the condition would be *that there is really a car there*. The equivalent for the 'moral hallucinator' such as the lynch mob member would be *that there really has been an act such that it is proper for me to hang its perpetrator from the nearest tree*, or some such. Assuming that the lynch mob member is not mistaken on some matter of material fact, how could the mental content expressed by *such that it is proper* be shown to be, or not to be, a moral hallucination? The perception hallucinator knows

how the world would be different if his perception was wrong; what could the moral hallucinator know?

The pure retributivist could make an appeal to the power of consensus, especially as embodied in good ethical thought. As Moore himself has said, "we 'see' that an action is wrong by applying the best moral theory we have about wrongfulness to the action before us."[15] Few will deny that Alyosha judges the nobleman's action as wrong by applying a good moral theory, namely that cruelty is wrong. But the pure retributivist needs more than this to justify his position. The conviction that Alyosha's 'Shoot him!' intuition is informative about moral reality depends on two further claims: (a) that the general theory that our moral intuitions are attributable to the moral properties of actions is true (so that a correctly reported, generally-shared, intuition is good evidence of a moral property); and (b) that this *particular* retributivist intuition, as glossed by Moore (that is, that punishment-requiring desert inheres in the nobleman's action) is in fact generally shared and is an example of such derivation of an intuition from the moral properties of an action.

There is more than one candidate, however, for a general theory about moral intuitions, and there may be more than one way of analysing a particular intuition. The appeal to the moral intuitions of the well-adjusted, that standard move of eighteenth-century moral sense theory, loses much of its force if different people supposed to be guided by their moral intuitions view things differently. Plausible alternative intuitions on a specific question, such as those I will mention later in connection with the Alyosha response, therefore tell against any claim that appeals to

[15] Moore 1982, 1133.

an innate moral sense. They force us either to conclude that some people (which ones?) have a defective moral sense, or to doubt that the moral sense exists, or tells us much that is reliable. Moore is aware of the objection and has appealed to a coherence criterion, modelled on the ontology of the natural sciences, to eliminate these problems. "To make the positive case for the existence of moral entities and qualities would be to itemize those items of our experience for which the best explanation would be the realist one: [that] we have such experiences because there are moral qualities."[16] The implication is that this realist explanation of our most compelling emotional responses to crime –that these are caused by moral facts of the matter, notably by the reality of the phenomenon of personal, moral, punishment-requiring desert- will fairly obviously knock out any rival explanations. The strength and near-universality of the Alyosha intuition then makes it, for Moore, one of the most reliable building bricks of a general moral theory of punishment.

It is striking, however, that even writers who accept a moral realist picture do not necessarily report the retributive intuition claimed by Moore. On the contrary, they are as divided as the rest of us when the specific question arises of the justification for afflicting wrongdoers. For example, in the pioneers of 'moral sense' theory, Hutcheson and Shaftesbury, the retributive intuition founded on punishment-requiring desert is conspicuously absent, though it does make an appearance before long with Butler and Adam Smith.[17] Hutcheson in particular draws an important distinction between the immediate anger we feel towards an evildoer, and the feelings that ensue when we reflect "calmly", or "with a sedate temper". In the latter frame of mind, we

[16] Moore 1982, 1124.

[17] See Butler, 'Dissertation Upon the Nature of Virtue' (Butler, 1967), 249-250; Smith 1982, 112-113.

conclude that "no Misery is farther the Occasion of *Joy*, than as it is necessary to some prepollent Happiness in the Whole".[18] The implication that the law should reflect the more 'sedate' judgment, and that it is this, not the retributive passion, that represents the moral sense at work, is clear. Butler's position is closer to Moore's, though with characteristic caution he remarks that our "sense of [actions] as of good or ill desert… may be difficult to explain".[19] Kant himself recognises the difficulty of explaining why moral wrong should in itself require the wrongdoer to be afflicted, treating this rather as a presupposition of punishment than as a belief in need of justification.[20]

Some facts of the matter, moreover, are less controversial than the existence of punishment-requiring desert, and other theories of the relationship between emotions and morality may prefer to build on these. The 'Alyosha' response suggests, for example, various conclusions about human nature. It suggests that most people feel a special protective care towards small children and are especially grieved by harm to them. Since the actions of a person motivated by this emotion will in general be socially useful, a philosopher such as Hume would count it among those admirable moral sentiments that move us to 'give a preference to the useful above the pernicious

[18] Hutcheson 2002, 58-59. Hutcheson's references to punishment are resolutely proto-utilitarian., e.g. 198 ("Human *Punishments* are only *Methods of Self-Defence*; in which the *Degrees of Guilt* are not the proper Measure, but the *Necessity of restraining Actions for the Safety of the Publick*"). For Shaftesbury's view, see Shaftesbury 2001, 36-37.

[19] Butler 249.

[20] "In punishments, a physical evil is coupled to moral badness. That this link is a necessary one, and physical evil a direct consequence of moral badness, or that the latter consists in a *malum physicum, quod moraliter necessarium est* [a physical evil that is morally necessary], cannot be discerned though reason, nor proved either, and yet it is contained in the concept of punishment, that it is an immediately necessary consequence of breaking the law…." (Kant 1997, 308).

tendencies'.[21] For Hume, a sentiment is accredited as a moral sentiment by the usefulness of the actions it typically motivates, and the approbation with which society views those actions. The Alyosha response also tells us that most people get very angry about cruel acts against innocents, and want to retaliate against the offender. It is not quite so uncontroversial that acting on this emotion will be socially useful, but a strong case can be made out for the view that, on balance, such indignation motivates courses of action that are beneficial.[22] The moral role assigned to emotions in this Humean analysis is instrumental, not epistemic: emotions are good if they motivate beneficent actions, bad if they motivate harmful ones, but the direction of such emotions is not supposed to provide evidence about the moral features of the world. Rather, a prior moral principle of utility assigns the emotions their positive or negative moral status.

Moore could object that this Humean analysis in effect begs the question, by failing even to consider the possibility that the Alyosha response provides evidence, not merely about the emotional life of human beings, of which it is an example, but about the moral world of human actions which is its intentional object. Again, however, rival theorists of punishment would be capable of offering alternative general theories that met this objection. These theories might accept moral realism, but identify a different way in which moral qualities inhere in or supervene upon human actions,[23]

[21] Hume 1998, 158.

[22] Butler, 'Upon Resentment' (1967, 130-131) argues that some resentment against offenders is necessary to counteract the effect of compassion, which might otherwise prevent us from punishing them at all.

[23] An example would be Schopenhauer who, in *On the Basis of Morality*, appeals to our intuitions with a number of examples in his attempt to convince us that "boundless compassion for all living things", proceeding inevitably from an innate tendency of character, is the single foundation of pure moral conduct (Schopenhauer 1995, esp. s. 19, 167-187). For his correspondingly anti-retributivist justification of punishment, see Schopenhauer 1960, 102-103.

or might propose a quite different vision of the relation between our experiences and our moral judgements.

One theory, which avoids making any ontological claims about moral entities and qualities, or postulating a distinct moral sense, runs as follows. Human beings possess an empirically demonstrable *moral faculty*. A faculty, as I use the term here,[24] is 'an ability or aptitude... for any special kind of action':[25] a capability which all human beings possess in a rudimentary form, but which is susceptible of development to a high level of self-awareness, complexity and codification, both within an individual life and within the life-history of the human species. An example of such a faculty would be *technology*: the faculty of using tools to realise our aims. A plausible candidate for identification as the moral faculty would be our capability of altruism or practical benevolence, of rationally pursuing the interests of others, in parallel with or in preference to our own. Our emotions, on this theory, are engaged by morally significant actions and events because the interests of others are often important to us emotionally, as well as morally. Even if the people we personally care for are not directly affected, we can often vividly imagine them in the place of those who are affected. Moreover, once we have formed moral codes, principles and practices that we believe are the best way to realise the aims of altruism, we tend to become emotionally committed to them, and to dislike those who reject or disobey them. In these considerations, especially the first, we can find both an explanation of the Alyosha response, and a reason for caution about the injurious actions to which it might impel us.

---

[24] As distinct from a common eighteenth- and nineteenth- century usage in which 'moral faculty' is merely a synonym for 'moral sense'.

[25] *Oxford English Dictionary*, definition I.1.a.

A retributivist might argue that we should identify the moral faculty instead with the capability of 'doing justice, creating so far as possible a world in which virtue is exactly correlated with happiness and vice with misery'. Thus, until one conception knocks out the other, we have two candidate 'moral faculties', each unquestionably at work in the world and undergoing continual development and codification. These conceptions would impel us towards different views on the justification of punishment, easily recognisable as the competing options in the present debate. Each could assign a valuable role to a set of emotions, both in motivating the realisation of the faculty and in providing touchstones for success. They would be slightly different sets of emotions, but not so different that either set could credibly be said to be the preserve only of an emotionally ill-adjusted minority. The Alyosha response would be met with reservation by the altruists, with endorsement by the devotees of justice. If these various alternatives are at all plausible, then Moore's invocation of a 'best moral theory' to assign moral significance to emotional responses gets us little further forward. There are many such theories. Moore needs to show –so far as this part of his argument for retributivism is concerned- that the power of the Alyosha response is such that the theory he constructs around it must be the right one.

As noted already, a quality-control criterion suggested by Moore is that if the possession of an emotion makes us more virtuous, this is "a good heuristic for coming to moral judgements that are true".[26] That we should perceive a broad correlation between virtue-enhancing emotions and good moral judgements is not surprising, since anyone who reflects seriously on some moral question is likely to find herself in

[26] Moore 1997, 134. Moore accepts that this criterion is "not infallible", since virtuous feelings can lead us to blame ourselves wrongly, and true judgements (such as that each person deserves the fruits of her own labour) can be prompted by emotions that are less than admirable.

a virtuous, or at least not frivolous, mood, and someone in this frame of mind is also more likely to make a decent shot at good moral judgement than someone in the grip of vicious emotions. It does not follow from the frequent co-occurrence of these two things, the virtuous emotions and the valid moral judgement, that the former provides evidence to support the latter. Whether the "Shoot him!" intuition of Alyosha arises from an emotion conducive to virtue is debateable: if we are sure that it is, that is likely to be because we already are confident that there is a morally valid judgement underpinning it.

At best, then, Moore's moral realist explanation of the evidential significance of emotions, considered purely as a structure of argument, fails to knock out other candidates for the best moral theory. It might still become the prime candidate, however, if the emotional responses he cites, and seeks to provoke in his readers, were sufficient to sweep us away with their intuitive power, so that we were forced to accept that the actions they suggest must be built into any acceptable theory of punishment. It isn't entirely implausible that some emotions might have this compelling power. Many people believe, for example, that the near-universal horror and repugnance at the idea of torturing children is sufficient to ensure that any moral theory that fails to endorse it, and to incorporate its implied deontological imperative at a foundational level, must be rejected.

But if we now turn to Moore's specific claim about the relevance of the Alyosha response, we find a further difficulty. We have agreed that most people would echo Alyosha's "Shoot him!" But the near-universality of this kind of response to truly outrageous acts of cruelty may obscure some important distinctions. Moore's claim is

not that the nobleman should be (lawlessly) shot on the spot, in expression of the outrage of an immediate observer -which is the form an unreflective satisfaction of the retaliatory intuition might well take. Moore's claim must be that the emotion provides evidence of a moral reality which justifies a pure-retributivist conception of criminal law. Yet an allusion to legal process may not be so nearly universal a component of the intuition.

This distinction between lawless and lawful action is important, not least because the creation of the process of law is itself an event of moral significance. Civilised society is not just an orderly arrangement of the emotional dynamics of the lawless world. The emotion (E) triggered by contemplating acts of violence as they unfold without the intervention of law may be different from that (E1) evoked by contemplating the same events in the context of a legal process. The pure retributivist might argue that our emotions developed in the state of nature reveal an objective moral reality, and that it is the message of those emotions on which in the world of law we ought to act.[27] But if it is indeed true that we have emotion E in the lawless world, and emotion E1 in the world of law, Moore needs to explain why emotion E accesses an objective truth whereas E1 doesn't. The truths of the lawless world, accessed by the emotions of the lawless world, would need to be justified as the determinant of our actions in the world of law. I'm not sure how Moore would defend this, and both a general and a particular objection to this move may be suggested. The general objection is that, in creating law-governed societies, we create a new moral landscape, properly generating different emotions: much as, in creating the institution of marriage, we

[27] This might be likened to Locke's "strange doctrine" that our right to punish as a society is derived from a natural right each of us possesses in the pre-social state (Locke, 1988, ss. 8-9, 272-273).

create the possibility of new or at least modified kinds of erotic sensibility. *Actions* which may be morally defensible in order to establish law in a lawless society (for example, killing an enemy in civil conflict) may become indefensible once law has been established; and if appropriate actions can be different as between lawless and lawful worlds, so can appropriate emotions. The particular objection is even simpler: the prisoner in a criminal trial is in a situation wholly different from that of an agent in the midst of lawless conflict, being at our mercy. It is quite natural that the emotions we feel about him in that situation may be different from those we feel as he commits his act against a helpless victim. This is not to deny that, even in the calm of a courtroom, violent retaliatory emotions may be felt by a victim, a victim's survivors, or others who empathise with them. But Moore's claim is not that we should be guided by the emotions of victims, but that we should be guided by those of well-adjusted observers.

Moore is, I believe, right to reject the rationalisation that explains our urge to do violence to the sadistic nobleman as motivated by considerations of deterrence or reform. I am less convinced, however, that he is right when he includes incapacitation in his list of alleged "bad reasons for what we believe on instinct anyway". [28] Certainly our response to the Dostoevsky example is not fully captured by the rational calculation, "If we remove this man permanently from the scene, he will not be able to carry out similar crimes again in the future". That analysis in terms of a future benefit would ignore our emotional focus on the terrible deed that has just been committed. It is rather that part of our anger (I speak to my own intuition here) may be a desire, not so much that the nobleman should suffer pain or reproof, as that he

[28] Moore 1997, 99.

should be obliterated, as if his act had qualified him for removal from the human race. A colleague once half-jokingly asked me if there existed an 'eradicative theory of punishment': a theory which holds that the purpose of punishment is to remove from the world those who spoil it for others by their intolerable acts. The real-life psychopaths and sadists whose crimes Moore recalls would be candidates for such eradication.

There has not, as it happens, arisen an 'eradicative theory', because when we are designing institutions of law and punishment, we think beyond the emotions appropriate to a world of lawless conflict. When we read Moore's narratives of atrocious crimes, our emotions do not situate us in a courtroom or even a prison, but at the moment of the atrocious action itself, when the powers of law are impotent. Who would not want to shoot on the spot the Russian nobleman, or the concentration camp guard, still wandering the ruins of a partially liberated camp, who has just needlessly murdered a surviving prisoner? And who would not respect such an action, even though the law might forbid it? We would respect it because it is the kind of action that we can imagine ourselves, whom we think of as morally serious and reflective persons, nevertheless being moved to perform by intolerable rage and indignation. The emotions that arise when we imagine ourselves into these situations of lawless conflict cannot be equated with those that arise when the offender is in the courtroom or the prison cell, at the mercy of the law. Probably the mass of human beings would indeed be better off if a small minority of grossly anti-social individuals were eradicated: certainly plenty of people feel this. But in criminal justice, the public authorities have responsibility for the fate of the 'world' that is constituted by the individual offender's experience; the person who would not have hesitated to kill in a

moment of lawless conflict might, in that situation, pause for reflection over the purposes that are to be served by afflicting the offender. The pure retributivist still needs to explain to this hesitating observer why the appalling acts of an individual make the creation of suffering in that individual's world a good thing to do, rather than a bad thing to do that can only be compensated by the prospective reduction of suffering elsewhere.[29]

Moore works hard to clear the retributive emotions he endorses from the accusations levelled against them by liberal humanitarians on the one hand, and by Nietzsche on the other –accusations that they arise from sadism, fear, mob emotion, resentment arising from weakness, a cowardly desire to retaliate against a defenceless enemy, etc. Such irrational, confused or malicious emotions undoubtedly exist, but Moore is quite right to point out that the outrage we feel when X afflicts Y for no good reason, far from being malicious, is founded in a morally commendable attitude, namely our fellow-feeling for Y. But what should happen to our fellow-feeling for X? Moore needs to show that certain emotions (emotions that can be cleared of malice or other kinds of inappropriateness) provide evidence for the necessary sequel of X's moral

[29] Ten, following Kleinig, argues that the fact that we can contemplate with equanimity that suffering which consists in the disappointment of a morally bad desire (for example, the suffering of a greedy person who fails to get more than his fair share, or a Nazi who never gets a chance to persecute Jews) calls in question the intuitive plausibility of the principle that "suffering is bad, no matter whose suffering it is" (Ten 1987, 47-48; Kleinig 1973, 67). I do not believe these supposed counter-examples erode the principle that suffering is bad. To accept that the suffering of the Nazi is bad (it may indeed have a special additional kind of badness for him, that of *undergoing* a pathological desire, quite apart from its frustration) is not to suppose that what he desires to do is good, or other than very bad. We are *glad* that he suffers the deprivation because –though he does not agree with this- the deprivation of a possibility that would be harmful to others is good, and so is the discouragement given by frustration to further attempts; not because it is intrinsically good that he suffer. The priority of this gladness over any compassion we might feel for a person undergoing a frustrated desire is more than sufficient to account for Ten's 'equanimity'. We may, of course, in reality be gratified by the thought of the Nazi's suffering. Such feelings are understandable and may be of instrumental value in motivating people to combat immoral attempts, but we do not need to claim moral credit for gratification in suffering. (Macaulay famously said that the Puritans hated bear-baiting not because it gave pain to the bear but because it gave pleasure to the spectators. This would be no discredit to the Puritans as long as their objection to the pleasure was that it would plausibly motivate further acts of bear-baiting.)

responsibility for crime in X's suffering punishment, and that this way of treating X is consistent with an acceptable kind of fellow-feeling towards him as a human being. Moore tries to show this by appealing to an introspective intuition.

> [A]sk yourself: What would you feel like if it was you who had intentionally smashed open the skull of a 23-year-old woman with a claw hammer while she was asleep, a woman whose fatal defect was a desire to free herself from your too clinging embrace? My own response, I hope, would be that I would feel guilty unto death. I could not imagine any suffering that could be imposed upon me that would be unfair because it exceeded what I deserved….[30]

Moore rejects, rightly in my view, the idea that this self-damning intuition can be resolved into the thought that we should somehow make compensation for our crime. "Corrective actions do not satisfy guilt." He is also properly severe on brutal criminals who, like the killer Richard Herrin to whom his example alludes, reconcile themselves too readily to their guilt and soon begin to complain that their sentences are excessive. After dismissing these attempts to deny or attenuate appropriate guilt, Moore concludes, on our behalf and his, if not Richard Herrin's, that

> Our feelings of guilt… generate a judgement 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 judgement is correct, generated as it is by emotions whose epistemic import is not in question.[31]

[30] Moore 1997, 145.

[31] Moore 1997, 148.

We can agree that the epistemic content of the imagined murderer's emotions is sound. He feels terrible because the deed he has done is a dreadful one, and the name for such emotions is guilt. He is not deluded, so they are appropriate emotions. To recognise one's guilt is indeed virtuous, since it shows awareness of which acts are moral and which are immoral, and a willingness to imagine vividly the consequences of one's actions; and to possess these dispositions is obviously more commendable than to lack them. But the claim that these first-person emotions "generate a judgement that we deserve the suffering that is punishment" is open to an equivalent objection to that which has already been levelled against the claim that the third-person emotions of horror and indignation generate "an intuitive judgement that punishment… is warranted".[32] Just as the retaliatory emotion we feel in response to another's atrocious crime is not, or at least need not be, framed within the discourse of lawful punishment, so the self-damning intuition of the person guilty of such an offence need not involve any such framing.

Many people who have committed crimes such as Herrin's have then taken their own lives. Others may have hoped to die but been unable to summon the resolution to commit suicide. We can all understand the emotions that might motivate such responses. But neither suicide nor the wish to die reflects the intuitive judgement Moore needs: "*I deserve lawfully imposed suffering*- not just in the sense that I know that there is a severe penalty for murder and that my act makes me eligible for it, not just in the sense that having unjustifiably killed someone I have broken the social compact and so have no ground to complain if I am killed, not just in the sense that it

[32] Moore 1997, 99.

is for the general good that the law should forbid and penalise murder, but - *as a necessary sequel to my murderous act, irrespective of any other considerations*". I do not think this fairly specialised judgement need be generated at all. Why should the self-condemner think of the law at all? Law is not everywhere.

The pure retributivist might reply that the killer's 'guilty unto death' feelings are sufficient to tell us that, in the judgement of the person we propose to punish, suffering or death is an appropriate sequel: criminal justice, then, in our world which happens to be law-governed, provides the method for delivering it. But this reply assumes (as in the parallel case of the emotions of the well-adjusted observer) that the emotions of the lawless world (in this case, those of the killer) should determine, or at least circumscribe, the actions of the law-governed one. It is not clear why we should accept this, given the objections stated above. Even a murderer with the most virtuous of remorseful emotions might form different, less wholly self-destructive, judgements if allowed to reflect in the context of the legal process. He might yearn to live a life of self-effacing service to others, for example: to live with his guilt as a motive for better actions, and not die from it; to suffer for some purpose. The forward-looking aspect to this yearning is not evidently less virtuous than the emotions identified by Moore.

So far I have assumed that Moore's account of the guilty-but-remorseful murderer's initial intuition is one we can all accept. However, a more likely paraphrase, in my view, is "I wish to leave as quickly as possible this world that will for ever be dominated for me by the memory of a horrible act, and the thought of its terrible consequences for my victim and others, with no prospect of peace of mind for me again". Moore might protest that this is a different, and essentially selfish, view, not

the true judgement of guilt, because it fails to focus specifically on the intrinsic justice of lawful retribution. But to make this objection would be to insist on defining a 'true judgement of guilt' in a way which makes it true by definition that it entails pure retributivism, while at the same time making it far less plausible to claim that our shared intuitions will support it.

In any case, even if (as I do not believe to be correct) the true judgement of guilt took precisely the form Moore needs, the legal-punishment-seeking form, it would still not follow that the state should act in accordance with that judgement. Moore might claim that the imagined murder's legal-punishment-seeking judgement of guilt is uniquely rational, and so should be binding on all others. But this would be implausible. Whatever sequel the self-judgement of a murderer in the immediate aftermath of a crime might propose, it is possible that judgements by others would lead to different conclusions: for example to the conclusion that a murderer's self-condemning 'guilt unto death' should be refused, in favour of requiring him to make reparation, for the rest of his natural life, for the harm caused by his action, and to reform his character if possible. This would be no less a form of punishment, but it would not correspond to the judgement of guilt as Moore imagines it; and it would be a consequentialist project. Though it might be abused, it is not self-evidently less rational than complying with the 'guilt unto death' intuition.[33]

It could be argued that the offender's legal-punishment-seeking true judgement of guilt provides a binding rule for all of us, because such a rule alone can be willed to be a universal maxim: *Always act in such a way that the author of every crime*

[33] And, to judge from *Crime and Punishment*, it is what Dostoevsky would have believed.

receives due punishment at the hands of the law, perhaps. But a consequentialist could accept an interpretation of this maxim, in which 'due punishment' means 'the punishment that for various consequentialist reasons, including the social benefit of consistency in punishing crimes, it is right to give'. The pure retributivist would have to rephrase the maxim to eliminate such interpretations: *Always act in such a way that the author of every crime receives due punishment at the hands of the law, for intrinsic reasons and not consequentialist ones.* Then the task of explaining the force of the intrinsic justifications for inflicting suffering, without drifting into consequentialism, would bring us back to the gap. And this revised maxim is not self-evidently the sole candidate for a maxim about punishment that could be willed to be universally believed and acted upon. Here is another: *Establish institutions of punishment on such a basis that the good of all citizens, and the avoidance of their harm, is most nearly achieved.* The good of all might, of course, best be achieved by distributing punishment on the basis of desert, as Rawls, Hart and others maintain.[34] But this thought brings us back to the multiple senses of 'desert', discussed earlier. Deserving punishment need not –and in the accounts of Rawls and Hart does not- have Moore's sense, of an intrinsic personal moral desert which provides both necessary and sufficient conditions for punishment. It need only signify that those punished are, in fact, judged to have intentionally committed the crimes to which the punishments are attached by law.

Acknowledgements

I am grateful to my colleagues Terry Hopton and And Rosta, and to a number of anonymous referees, for invaluable criticism, discussion and encouragement

[34] Rawls 2001, 20-29; Hart 1968, 8-24.

WORKS CITED

Butler, Joseph, 'Upon Resentment', in *Fifteen Sermons*, ed. W. R. Matthews (London: G. Bell & Sons, 1967), 120-132.

Duff, R.A., 'Penal Communications: Recent Work in the Philosophy of Punishment', *Crime and Justice* 20 (1996), 1-97.

Hart, H.L.A., *Punishment and Responsibility* (Oxford: Oxford University Press, 1968), 1-27.

Hume, David, 'Concerning Moral Sentiment', *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp (Oxford: Oxford University Press, 1998).

Hutcheson, Francis, *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. A. Garrett (Indianapolis: Liberty Fund, Inc., 2002).

Kahan, D. M. and M. C. Nussbaum, 'Two Conceptions of Emotion in Criminal Law', *Columbia Law Review* 96 (1996), 270-374.

Kant, I, *Lectures on Ethics*, translated by Peter Heath, ed. Peter Heath and J.B. Schneewind (Cambridge: Cambridge University Press, 1997).

Kleinig, John, *Punishment and Desert* (The Hague: Martinus Nijhoff, 1973).

Locke, John, *Second Treatise of Government*, ed. P. Laslett (Cambridge: Cambridge University Press, 1988).

Matravers, Matt, *Justice and Punishment: The Rationale of Coercion* (New York: Oxford University Press, 2000).

Moore, Michael S., 'Moral Reality', 1982 *Wisconsin Law Review* (1982), 1061-1156.

Moore, Michael S., *Placing Blame* (Oxford: Clarendon Press, 1997).

Murphy, Jeffrie and Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988).

Rawls, John, 'Two Concepts of Rules', *Collected Papers*, ed. Samuel Freeman (Cambridge, Mass: Harvard University Press, 2001), 20-47.

Schopenhauer, Arthur, *Essay on the Freedom of the Will*, translated by Konstantin Kolenda (Indianapolis: Bobbs-Merrill, 1960).

Schopenhauer, Arthur, *On the Basis of Morality*, translated by E. F. J. Payne (Oxford: Berghahn Books, 1995).

Searle, John R., *Intentionality* (Cambridge: Cambridge University Press, 1983).

Shaftesbury (Antony Ashley Cooper, Earl of Shaftesbury), 'An Inquiry Concerning Virtue and Merit', *Characteristicks of Men, Manners, Opinions, Times*, Vol. 2, ed. Douglas Den Uyl (Indianapolis: Liberty Fund, Inc., 2001), 3-44.

Smith, Adam, *The Theory of Moral Sentiments*, ed. D. D. Raphael and A.L. Macfie (Indianapolis: Liberty Fund, Inc., 1982).

Ten, C.L., *Crime, Guilt and Punishment* (New York: Oxford University Press, 1987).