

A NOTION OF MORAL RESPONSIBILITY IMMUNE TO THE THREAT FROM CAUSAL DETERMINATION

Derk Pereboom, Cornell University

Forthcoming in *The Nature of Moral Responsibility*, Randolph Clarke, Michael McKenna and Angela Smith, eds., Oxford University Press, 2015.

Penultimate draft

Abstract: This article sets out a notion of moral responsibility that incorporates the central features of the answerability conception advocated by T. M Scanlon, Hilary Bok, and Angela Smith, and of Michael McKenna's more specific conversational account, but which excludes any notion of desert, whether basic or non-basic. The point of blaming and praising on this notion largely forward-looking: its main objectives are protection, reconciliation, and moral formation. Agents are blameworthy and praiseworthy by virtue of being appropriate recipients of blame and praise given these aims. Blaming on this conception can involve causing harm, but the justifiability of such harming does not reintroduce the legitimacy of desert. The resulting notion of moral responsibility is immune to any threat from the causal determination of action.

Keywords: determinism. moral responsibility, blame, praise, desert, moral formation, protection, reconciliation, moral obligation, deontology.

Our practice of holding each other morally responsible involves a number of different senses of moral responsibility. It's generally agreed that this practice features the

notion of desert. In the basic form of desert, someone who has done wrong for bad reasons deserves to be blamed and perhaps punished just because he has done wrong for those reasons, and someone who has performed a morally exemplary action for good reasons deserves credit, praise, and perhaps reward just because she has performed that action for those reasons (Feinberg 1970; Pereboom 2001, 2014; Scanlon 2013). This backward-looking sense is closely linked with the reactive attitudes of indignation, moral resentment, and guilt, and on the positive side, with gratitude (Strawson 1962); arguably because these attitudes presuppose that their targets are morally responsible in the basic desert sense. Basic desert responsibility is prominent in the philosophical discussion of free will because it is the sense most clearly threatened by causally deterministic histories of action that trace back beyond the agent's control, and it is salient in everyday life because the emotions connected with it are especially vivid.

There may be senses of moral responsibility that involve a non-basic variety of desert. Essentially forward-looking notions of holding agents deserving of blame and punishment have been defended on consequentialist or contractualist (Dennett 1984, 2003; Lenman 2006; Vargas 2007, 2013; Vilhauer 2013) grounds. But there are other senses of moral responsibility, also part of our practice, that do not involve a notion of desert, and are not threatened by causal determination of action. Some of these senses are distinctively forward-looking. On Moritz Schlick's (1937) proposal, the point of blaming and praising is to reduce the incidence of bad action and to increase the frequency of good action. As P. F. Strawson pointed out, such a forward-looking "optimist" conception does not capture our entire actual practice of holding morally responsible, and in particular misses the component connected with the

reactive attitudes. But it is nonetheless an important element of that practice. When parents and teachers blame children for bad behavior, it's often this forward-looking sense that's in play.

There are in addition further senses of moral responsibility that do not involve desert but are nonetheless backward-looking. On George Sher's account, blame is essentially a certain belief-desire pair: the belief that the agent has acted badly or that he has a bad character, and the desire that he not have performed his bad act or not have the bad character (Sher 2006: 112). In his characterization, blame neither essentially involves negative reactive attitudes, nor is it forward-looking. Even though Sher's notion is backward-looking, it is also not at issue in the free will debate, since it is evidently not threatened by causal determination. A related backward-looking sense that is arguably also not at play in the free will debate is the notion of normative moral judgment that Angela Smith (2008) invokes. When we judge that Zoë's decision to evade taxes was wrong, we are making a normative moral judgment, and given the appropriate context, this can qualify as a way of holding her morally responsible. Yet it is not one that is threatened by causal determination -- or at least not clearly so (more on this later). For an analogy, suppose Yuri made an error in balancing his checkbook. When we say his calculation was wrong, we are also making a normative judgment, but not one that is threatened by his having been causally determined by factors beyond his control in making his calculation.

Contrary to what is at times supposed in the philosophical discussion, it's mistaken to claim that the term 'moral responsibility' has a single sense. Often terms that have a long history of use have a number of senses. Even if a referring term with such a history originally had just one specific referent, over time it is apt to be applied to similar but

distinct referents, thereby acquiring different senses. The terms 'moral responsibility' and 'blame' plausibly have this profile. Ludwig Wittgenstein (1953) famously makes the general point with the example of 'game.' There is a family resemblance among the referents of this word, but these referents do not exhibit a simple and non-disjunctive common characteristic, and the word may therefore have multiple senses.

There are reasons to be skeptical of any notion of moral responsibility that involves desert. One worry for the basic desert sense is that for an agent to basically deserve a harmful response she must have a kind of free will that is unavailable to us, and the free will skeptic contends that this concern can't be successfully countered (Galen Strawson 1986; Waller 1990, 2011; Pereboom 1995, 2001, 2014). As noted earlier, one might argue that some desert sense of moral responsibility can or should be retained because doing so stands to bring about good results, while not doing so would be deleterious (Dennett 1984, 2003; Vargas 2007, 2013). One might, for example, endorse a practice-focused consequentialism, according to which we are to compare the consequences of the competing candidates for moral practice to see which one comes out best. And it may be that the conception that retains our basic desert assumptions comes out best. But putting this conception into practice would involve our thinking and acting as if agents are morally responsible in a sense in which they are not, and thus it would involve acting as if false beliefs were true, and this would seem unfair (Waller 1990: 130-35). It would be *prima facie* preferable if we did not live with this illusion, and the position I will set out meets this objective.

A further concern for desert senses of moral responsibility is that for a number of contending general normative ethical theories the notion of desert is relegated to the role

of an awkward supplement. Any role for desert in typical consequentialist views is uncomfortable, and despite Kant's well-known (1791/1963) invocation of desert in justifying criminal punishment, that appeal seems unrelated to any formulation of the Categorical Imperative, which he held to be the supreme and comprehensive moral principle. Another worry is that negative desert, at least in its basic form, appears to involve the idea of harm as an intrinsic good, which seems dubious. Motivated by these concerns, I will propose a view that rejects all desert-involving senses of moral responsibility.

A notion of moral responsibility free from the threat of causal determination.

The free will skeptic Joseph Priestley (1788/1965) and his revisionary compatibilist cousins such as J. J. C. Smart (1962) contend that there is a forward-looking notion of moral responsibility that is immune to any threat from causal determination. On the type of view they propose, the justification and goal of the practice of blaming and praising is to weaken dispositions to misconduct and strengthen dispositions to good behavior. The dispositions addressed already exist and in paradigm examples are manifest in past actions. In the case of justified blame these dispositions are reasonably assumed to persist unless corrective measures are taken. Blaming of this kind, then, addresses past misconduct as a means to moderating or eliminating such a standing disposition. This conception need not advocate treating agents as if they are merely stimulus-response mechanisms. When an agent has acted badly, one might ask him: "Why did you decide to do that?" or "Do you think it was the right thing to do?" where part of the point of asking such questions is to communicate reasons to acknowledge and address a disposition to behave badly. If the reasons for his

behavior he provides in response to such questions confirm that he does indeed have such a disposition, it then becomes appropriate to request an effort to eliminate it (cf. Hieronymi 2001).¹ Engaging in such interactions will be legitimate in light of how they contribute to the agent's moral improvement. This model is a variety of the *answerability* sense of moral responsibility proposed by T. M. Scanlon (1998), Hilary Bok (1998), and Angela Smith (2008).

Michael McKenna (2012) has recently developed a conception of moral responsibility of this kind, one that turns out, with a few key revisions, to be amenable to the free will skeptic. In his *conversational* theory of moral responsibility, actions of a morally responsible agent are potential bearers of a type of meaning by indicating the quality of will that resulted in the action (2012, pp. 92-94; see also Arpaly 2006). My blaming an agent who manifests an immoral quality of will in action is an expression of an attitude such as moral resentment or indignation, and its function is to communicate to him my moral response to the indicated quality of will. Morally responsible agents understand that members of the moral community might attribute such a meaning to their actions. When actions are morally charged, they understand themselves to be introducing a meaningful contribution to such a conversational exchange. McKenna labels this initial stage of the conversation *moral contribution*. In the case of a prima facie immoral action, in

¹ Pamela Hieronymi (2001, p. 546) proposes that resentment is best understood as a protest; "resentment protests a past action that persists as a present threat." Resentment is not a feature of the forward-looking notion of blame I set out, but on this account a core function of blaming someone is to protest a past action of his that persists as a present threat, as in Hieronymi's view.

the second stage the agent is blamed by a respondent; McKenna calls this stage *moral address*. In the third stage, *moral account*, the blamed agent offers an excuse, a justification, or an apology. The respondent might at this point continue the conversation by forgiving or punishing the wrongdoer. In a subsequent stage the blamed agent may be restored to full status in the moral community. McKenna points out that not all blaming conforms to this model; blaming the dead, for instance, does not. Here he invokes a paradigm-similarity model for the meaning and extension of a concept (Rosch 1972, 1973). The blame conversation as he describes it is a paradigm case of blaming, and examples of other sorts, such as blaming the dead, qualify as instances of blaming because they are sufficiently similar to such paradigm cases.

I endorse such a model with a few modifications (Pereboom 2013ab, 2014).² I prefer not to characterize the first stage as appealing to notions as sophisticated as “introducing a meaningful contribution to such a conversational exchange with others.” Key to the first stage is that the agent at least implicitly think of her action as morally charged and as communicating a morally salient quality of will. In addition, when blame is at issue, in the first stage it will often be the case that an action that can reasonably be interpreted as immoral is performed. But this is not so for all instances of overt blaming, because we frequently blame due to mistaken perceptions or dubious second-hand reports of agents’ acting immorally, or to negligent or deliberate misinterpretations of actions, or to sheer fabrication. Often, then, it would be preferable for blaming not to precede a request for excuse, justification, or some other type of exonerating explanation. In such cases, if the

² McKenna is open to some of these amendments (in correspondence, cf. McKenna 2012, pp. 90-91).

target is overtly blamed before such a request, blaming will be morally faulty. If at the second stage it is open that the agent has an excuse or justification (and sometimes it isn't – when, for instance, the agent is clearly caught acting badly), a blaming response that communicates to her that she acted badly and without excuse or justification – that is, an accusation – will typically be misplaced. If the respondent makes an accusation at the second stage, and the agent offers an acceptable excuse or justification, an apology on the part of the respondent will be appropriate. If it was genuinely open that the agent had a valid excuse or justification but in fact had none, the prior accusation would have been inappropriate.

As I remarked earlier, the skeptic about desert can endorse an answerability notion of moral responsibility. She can also without inconsistency accept an amended version of McKenna's more specific proposal insofar as it avoids an appeal to basic desert and to expressions of reactive attitudes that are linked with beliefs about basic desert. McKenna argues, convincingly, that his general model is compatible with both endorsing and rejecting basic desert. I propose to ground this model for blame not in desert, but almost exclusively in certain forward-looking moral considerations. Some backward-looking elements in my view are required for justifying blame, for example that the agent did in fact act badly, but desert is excluded.

Specifically, I have in mind three non-desert involving moral desiderata: protection of potential victims, reconciliation in personal relationships and with the moral community, and formation of moral character (Pereboom 2013ab, 2014). Immoral actions are often harmful, and we have a right to protect ourselves and others from those who are disposed to behave harmfully. Such actions can also impair relationships (Scanlon 2009),

and we have a general moral stake in restoring impaired relationships through reconciliation. And indeed, we have a moral interest in the reformation of moral character afflicted by dispositions to misconduct. Agents are blameworthy and morally responsible, on this model, by virtue of being appropriate recipients of blaming that serves these aims. Blame justified in this way is largely forward-looking, because its aims are future protection, future reconciliation, and future moral formation. The immediate target of blame is often a past action, and in this respect such blaming will have a backward-looking aspect, but insofar as the purpose of blame is protection and moral formation, the past action will be addressed as a means to correct a persisting disposition to act immorally. To the extent that the objective of blame is reconciliation, the past action will also be addressed for its own sake.³

There may be instances in which an immoral action has been performed but without a persisting disposition so to act, or without a disposition that poses a genuine threat. Blame can then still have the point of reconciliation. There are also cases of persisting dispositions to act badly where the disposition has yet to be manifested in action. Such an agent can still legitimately be blamed for having the persisting disposition, and, more generally, on the largely forward-looking notion I advocate, persisting dispositions are what blame primarily addresses.

There is an account of praise that correlates with this conception of blame. Of the aforementioned goals of blaming – protection, reconciliation, and moral formation – the aim particularly amenable to praising is moral formation. We praise an agent for a morally

³ Thanks to Randolph Clarke and Dana Nelkin for drawing my attention to the backward-looking elements of this account.

exemplary action to encourage her to strengthen the disposition that produced it. This can also have a protective function, because strengthening such a disposition diminishes threats to others. Corresponding to reconciliation is the notion of recognizing successes and accomplishments in a relationship. Praising actions can have this role as well.

On this account, the agent's practical rationality is engaged in blaming and praising. In the case of blaming, at the stage of moral address we ask for an explanation of the action, and if it turns out that the agent has acted badly without excuse or justification, we aim for him to recognize that the disposition issuing in the action is best eliminated. In the standard case, this change is occasioned by the agent's recognition of moral reasons to make it, and part of the function of the moral conversation is to occasion awareness of such reasons (cf. Fischer and Ravizza 1998). In summary, then, it is the agent's responsiveness to reasons, our moral interest in protection, the agent's moral formation, and our reconciliation with him, together with the fact that he had indeed acted badly, that explains why he is an appropriate recipient of blaming. An analogous case can be made for praising.

The proposal as I've spelled it out so far makes no appeal to desert, whether basic or non-basic. Even though McKenna agrees that the conversational model need not appeal to basic desert, he nonetheless wants to make a case for a link between blame on this model and basic desert. He argues that this connection is illuminated by the following principle:

It is a noninstrumental good that, as a response to the meaning expressed in an agent's blameworthy act, that agent experiences the harms of others communicating in their altered patterns of interpersonal relations their moral demands, expectations, and disapproval. Because this is a noninstrumental good, it is permissible to blame one who is blameworthy. (McKenna 2012: 150)

The harms upon being blamed include the emotional pain of engagement in the moral conversation, and the alteration for the worse of personal relationships. McKenna argues that such harms can be construed as noninstrumental goods, and thus basically deserved. My sense is that the goods invoked turn out only to be instrumental to realizing other goods, such as the agent's membership in the moral community and his commitment to morality, and thus not basically deserved (Pereboom 2013ab, 2014). McKenna responds (in conversation) by contending that the harm of blame is partially constitutive of these noninstrumental goods, and thus not merely instrumentally valuable. I've replied that while it is plausible that certain uncontroversial examples of goods, such as mental and physical health, are partially constitutive of a noninstrumental good such as human flourishing, it is at least typically less credible that harms -- as harms -- are partially constitutive of noninstrumental goods and for this reason count as noninstrumental goods themselves. Vaccination may be required for physical health, while health is constitutive of flourishing, but it's not credible that the pain of vaccination is constitutive of flourishing, as opposed to merely being instrumentally required for it (2013ab, 2014).

In response to this objection, McKenna (in conversation) argues that in the case of grief upon the death of a friend, it's plausible that the pain of grief is a component of the good that it constitutes, by contrast with merely being instrumentally required for it. The closest analogy among the basic-desert presupposing reactive attitudes would be guilt. Perhaps it's also credible that the pain of guilt is a component of the good that guilt constitutes, and is not only instrumentally required for it. Thus McKenna's suggestion is that the pain of guilt, like the pain of grief, is a noninstrumental good due to being a component of a larger good.

Dana Nelkin (forthcoming) objects that although it may be that the aspects of blaming that McKenna identifies are noninstrumental goods, and that they are non-contingently related to harms, this does not show that these goods are good even partly in virtue of those harms. In addition, even if the harms are required for the good while not being instrumental goods, as McKenna proposes, that does not make them noninstrumental goods rather than not being goods at all. Applied to grief, supposing grief to be a noninstrumental good, and that the pain of grief is an essential component of grief, it may well be that this pain is not a good at all, but instead something bad required for a good.

Randolph Clarke (2013) suggests that we can get a sense for how the pain of guilt could be a noninstrumental good by reflecting on the idea that the guilty deserve to feel guilty at the right time and to the right degree. He proposes, first, that there is value in the recognition by an agent who is blameworthy that he is blameworthy. A further response, the feeling of guilt, would provide a morally valuable and intuitively fitting addition to this acknowledgement. This response would have value insofar as it is expressive of the agent's moral concern for the fact that he was the one who did the wrong and for those he has wronged. This insight serves as an avenue to seeing how the guilty could deserve to suffer the pain of guilt.

It is highly intuitive that there is value in the recognition by an agent who is blameworthy in the largely forward-looking sense that I have set out that he is blameworthy. This would, in my view, count as an acknowledgement of the truth. So far, desert is not implicated. But more significantly, I agree that a sense of pained remorse is a morally fitting additional response, without accepting that the guilty deserve to suffer such pain. Bruce Waller (1990) and Hilary Bok (1998) argue that the fittingness of such pain can

be accounted for by a recognition that one has not lived up to one's standards for morality and self-control, without the need to invoke desert. Waller describes a case in which one strikes a friend in a fit of anger, and subsequently feels deeply disappointed about one's deficient capacity for control. The disappointment is painful, and while the pain is appropriate it is intuitively not deserved (Waller 1990, pp. 165-66). Bok sets out an example in which one has done something wrong, on account of which one suffers a painful response, which she compares to heartbreak. (She calls this response 'guilt,' but we might prefer to reserve this term for the self-directed, basic-desert involving reactive attitude. We can substitute 'remorse' or 'regret' for Bok's 'guilt'):

The relation between the recognition that one has done something wrong and the guilt one suffers as a result... is like the relation between the recognition that one's relationship with someone one truly loves has collapsed and the pain of heartbreak. Heartbreak is not a pain one inflicts on oneself as a punishment for loss of love; it is not something we undergo because we deserve it... Similarly, the recognition that one has done something wrong causes pain. But this pain is not a form of suffering that we inflict on ourselves as a punishment but an entirely appropriate response to the recognition of what we have done, for two reasons. First, our standards define the kind of life we think we should lead and what we regard as valuable in the world, in our lives, and in the lives of others. They articulate what matters to us, and living by them is therefore by definition of concern to us. If we have indeed violated them, we have slighted what we take to be of value, disregarded principles we sincerely think we should live by, and failed to be the sorts of people we think we

should be. The knowledge that we have done these things must be painful to us.

(Bok 1998, pp. 168-69)

I think that Waller and Bok are right to contend that feeling pain on account of a recognition that one has not lived up to one's own moral standards or standards for self-control need not involve desert. By analogy, one might feel pained that on some occasion one failed to meet one's standards for chess or piano playing when one understands that one's substandard performance is due to factors beyond one's control, and that this pain not deserved.

Ben Vilhauer (2004) advocates an account of a pained response upon one's own wrongdoing that grounds it in sympathy with those one has wronged on which the feeling is fitting because the sympathy is morally appropriate. It's highly credible that such sympathy-based remorse is apt for motivating repentance and moral self-improvement, for reconciliation with those one has wronged, and subsequent restoration of one's integrity. Vilhauer contends that because such sympathy-based remorse is also other-directed rather than merely self-directed, it is morally preferable to remorse grounded in basic desert. Remorse on a basic desert conception has no essentially forward-looking moral objective. By contrast, sympathy-based remorse involves taking on the perspective of the agent one has wronged, which has morally beneficial consequences.

Intuitively, this can't be the complete story, for as Austin Duggan points out (in conversation), we find it highly plausible that a pained response to one's wrongdoing is appropriate and fitting for agents who are not disposed to feel pain upon the violation of their moral standards and standards for self-control, and who lack sympathy with those who have been wronged. But this intuitive sense can be accounted for by way of the largely

forward-looking goals of moral reform and reconciliation. For those who lack the predispositions that Waller, Bok, and Vilhauer describe, feeling pained upon having done wrong would nevertheless be valuable because it is liable to result in moral reform, and in reconciliation with those who have been wronged. In addition, we believe that it would be morally preferable on the basis our largely forward-looking grounds if they instead were disposed to feel pained upon the violation of their moral standards and to feel sympathy with those who have been wronged. Our sense of the fittingness of the pained response can partially be explained by this belief as well.

In summary, a painful response to one's own wrongdoing is intuitively appropriate. But the justification for the appropriateness of such a painful response need not appeal to desert. Instead, it might appeal to forward-looking considerations such as the moral reform of the wrongdoer, and sympathy and reconciliation with those who were wronged.

Blame, obligation, and wrongness

Against the forward-looking account of moral responsibility I've proposed one might object that for an agent to be blameworthy for performing action A even in this forward-looking sense requires that it she ought not have done A, and this in turn requires, by the 'ought' implies 'can' principle (OIC), that she could have refrained from doing A, and this is incompatible with the agent's causal determination by factors beyond her control. While my free will skepticism does not endorse the general causal determination of human action, it leaves it open as a serious possibility. According to the consequence argument causal determination is incompatible with being able to do otherwise, at least in one key sense, and on this understanding I find the argument compelling (van Inwagen 1983; Ginet

1966, 1990). Hence this objection poses a threat to my proposal. A second and related concern is that if moral obligation is undercut by the general causal determination of action, moral wrongness is also undermined. Ishtiyaque Haji argues that S has a moral obligation to perform A just in case it is morally wrong for S not to perform A, and, similarly, that S has a moral obligation not to perform A just in case it is morally wrong for S to perform A. Thus moral wrongness and moral obligation stand and fall together. I've claimed that Angela's Smith (2004) notion of moral judgment, which invokes the normative notion of moral wrongness in the case of immoral action, is not threatened by causal determination. But now it seems as if it is. Let's consider these two objections in turn.

A response to the first objection might take advantage the plausible claim that 'ought' has a range of correct uses, and as C. D. Broad (1952) suggests, it may be that not all are linked to an OIC requirement, or at least to an OIC requirement that facilitates the incompatibility of causal determinism and 'ought' judgments.⁴ First, Ruth Barcan Marcus (1966), Lloyd Humberstone (1971), and Gilbert Harman (1977) distinguish between an 'ought' that applies to action and one that applies to states of affairs.⁵ An 'ought to do,' Harman contends, "implies that an agent has a reason to perform an action, while an 'ought

⁴ I develop this line of response more thoroughly in Pereboom 2014, Chapter 6.

⁵ Cf., Kate Manne (2011). Mark Schroeder (2011) distinguishes between the action-related *deliberative* sense of 'ought', and the *evaluative* 'ought', as in 'Larry ought to win the lottery' where Larry has been subject to a series of undeserved misfortunes. Manne argues, plausibly to my mind, that it is important to see that the evaluative 'ought' applies not only to non-agential states of affairs, but also to actions.

to be' evaluates a state of affairs and does not by itself imply that any particular agent has a reason to contribute to bringing about that state of affairs" (Harman 1977: 87; cf., Humberstone 1971, Manne 2011). James Hobbs (2012) persuasively argues, certain 'ought to be' claims will have a loose kind of implication for reasons agents have to act. He proposes the following account: if X *ought to be* the case, then agents to whom the relevant considerations apply have a reason to act in ways that respect the value of X. But he contends, plausibly to my mind, that the satisfaction of this condition does not imply a route that is accessible to the agent to the realization of what ought to be. However, if instead an agent *ought to do* something, then she has a reason to do it, and there is such a route to what she ought to do. In a similar vein, Nelkin (2011: 111) contends that 'ought' propositions that specify what an agent ought to do are essentially action-directed, so that if 'S ought not do A' is true, then as a matter of the meaning of 'ought' judgments, or of the essential nature of obligation, S is thereby directed to refraining from A, and this entails that S can refrain from A (see also Copp 2008; Stapleton 2010 for similar views). One way to think about this distinction is that an 'ought to be' is an 'ought' of *axiological evaluation*, or sometimes of *axiological ideality*, which does not entail a 'can' claim, whereas an 'ought to do' expresses a demand made of an agent in a particular circumstance, which does entail that the agent can perform the specified action (cf., Humberstone 1971; Manne 2011). We can call this second type an 'ought' of *specific action demand*.

Hobbs's more precise compatibilist proposal for the ability entailed by an action-directed 'ought' is that if a person ought all-things-considered to do A, then she has the physical and mental ability, the skill, and the know-how needed for doing A, and she is in circumstances appropriate for doing A. But she need not have the motivation required to

do it. Accordingly, if an agent all-things-considered ought to do *A*, then she can do *A* in the sense that doing *A* is compatible with her abilities and her opportunities, but not necessarily with how she is in fact motivated. Significantly, in his view the general causal determination of action implies that doing otherwise is incompatible with the full range of causally relevant features of a situation. But that an agent ought to perform an action, all things considered, implies only that her performing this action is compatible with her abilities and opportunities, and not with the remaining crucial and causally relevant feature, her motivation. Thus, 'S ought not to do *A*' can be true in a situation in which *S* is causally determined by factors beyond his control to perform immoral act *A* if it's his motivation that's defective.

Hobbs tests his account with the objection that the relevant 'ought' claim in fact place a stronger demand upon an agent, one that it would be unfair to impose without being more sensitive to how an agent could be motivated as a matter of causal fact. He responds by arguing that it is unclear what this stronger sense of 'ought' would be: "the practical, action guiding 'ought to do' claims I have in mind very often express obligations, and the only sense in which 'ought' could express some stronger normative claim, one for which it might be unfair to fail to consider motivational obstacles, is one that is tied to blameworthiness, such that 'A ought to Φ ' more or less entails that *A* would be blameworthy if he failed to Φ . This is not a use of 'ought' that I am familiar with or that I find particularly useful" (Hobbs 2012). But I think it's plausible that if causal determination precludes alternative possibilities for motivation and thus for action in a sense relevant to the issues under consideration, then there is a core notion of 'ought to do' that will be undermined. If I know that in this relevant sense an agent could not have avoided lacking

the motivation required for refraining from performing an immoral action, it would be unfair, and I think, mistaken, for to claim that she ought not to have performed that action at that time.

But at the same time it might well not be mistaken or unfair for me to recommend to that agent that she perform an action of that type in the future, supposing that it's reasonable to believe that she has or will have the abilities and opportunity to perform the action, and that it's epistemically open that she will acquire the requisite motivation, and in particular if it's reasonable to believe that making this recommendation would contribute causally to bringing about the motivation. To recommend the action to her, I might tell her that she ought to perform the action at the future time, and do so appropriately and without making any kind of mistake. In my view, the sense of 'ought' invoked here would need to be distinct from the 'ought' of specific action demand. Accordingly, I propose that given determinism and that determinism precludes alternatives, when one tells an agent that he ought to refrain from performing an action of some type in the future, it's not the 'ought' of specific action demand, but rather the 'ought' of axiological evaluation that is legitimately invoked. This use of 'ought' proposes as morally valuable a state of affairs in which the agent refrains from performing the action and recommends that she not perform it. We might call this the 'ought' of *axiological recommendation*. Unlike the 'ought' of specific action demand, it is not an 'ought' of obligation. Supposing the general causal determination of action and that such determination rules out the relevant alternative possibilities for action, the use of the 'ought' of specific action demand could be correct and fair only if the agent is in fact causally determined to perform the action and one is reasonably sure that she is, which would typically be untrue. But under these

circumstances the 'ought' of axiological recommendation would not be similarly undermined.

Like the 'ought' of specific action demand, the 'ought' of axiological recommendation essentially concerns agents and actions they might perform. But as for all claims about what ought to be, this use of 'ought' should not be understood as presupposing a route accessible to an agent, by way of reasons for action, to her acting in a relevant way. One might be unclear about whether such a route is accessible, while the use of 'ought' is nevertheless appropriate.

How do these reflections bear on the relation between blameworthiness and 'could have done otherwise'? Blameworthiness in the forward-looking sense I've set out licenses the right sort of respondent to tell the agent who has acted badly that he ought not to act this way, where 'ought' has the sense of axiological recommendation, which in turn requires that it be epistemically open for the respondent that the agent will comply with this recommendation. Thus for an agent to be blameworthy for an action in the forward-looking sense it must be epistemically open that he refrain from performing actions of this type in the future. (But blameworthiness in this sense does not require that the agent ought not, in the sense of moral obligation, have done what he did.)

When is an action sufficiently bad to warrant blame in the forward-looking sense? It's natural to think that an action justifiably occasions this kind of blame when it's morally wrong for the agent to perform it. This brings us to our second objection, according to which judgments of normative wrongness are undermined by causal determination. Haji (1998, 1999, 2002) argues that due to the tight connection between moral obligation and

moral wrongness, the threat posed to judgments of moral obligation extends to those of moral wrongness. Crucial to his argument is the following principle:

S has a moral obligation to perform [not to perform] A if and only if it is morally wrong for S not to perform [to perform] A.

If this principle is true, then if judgments of moral obligation are ruled out by causal determination, judgments of moral wrongness will be excluded as well.

Although this biconditional principle may be attractive considered in the abstract, my sense is that it captures only one aspect of the complex notion of moral wrongness. This can be made plausible by reflecting on the fact that while the left-to-right half of this biconditional

If S has a moral obligation to perform [not to perform] A then it is morally wrong for S not to perform [to perform] A,

is evidently true, the right-to-left half, i.e.,

If it is morally wrong for S not to perform [to perform] A, then S has a moral obligation to perform [not to perform] A,

is less clearly secure. There's likely no example in which it's at all credible that an agent has a moral obligation not to perform an action while it is not morally wrong for him to perform it.

But there are cases in which it's intuitive that performing an action would be morally wrong for an agent, while it's less plausible that he has a moral obligation not to perform it. Imagine that a serial killer could not have avoided intending to kill due to his psychological disorder.

OIC provides an intuitive basis for denying that he was morally obligated to refrain from killing, while it remains evident that his intention was morally wrong (for opposing views, see Haji 2002 and Stapleton 2010).

An alternative non-deontological notion of moral wrongness, one that isn't biconditionally linked to moral obligation, accommodates this intuition. The core of Alastair Norcross's (2006) proposal for a purely axiological ethics involves specifying for each action-relevant situation the options for acting ranked in order of value realized. Value would include well being, but potentially also factors such as respecting of rights and absences of rights violations. An option for acting might then be counted as morally wrong when its value is low enough in the ranking for it to be morally justified for a relevantly positioned interlocutor to blame the agent for the forward-looking reasons I've set out. This proposal does not characterize wrongness independently of when it is appropriate to blame for these reasons, and thus it cannot ground the appropriateness of blaming in wrongness. Consequently, it does not accommodate one significant intuition we have about the relation between wrongness and the appropriateness of blame. But the proposal does have the consequence that it's appropriate to blame an agent only when his acting for some reason would be morally wrong, and in thus it satisfies another important intuition we have about this relationship.

Final words

I've set out a notion of moral responsibility that incorporates the central features of the answerability conception advocated by Scanlon, Bok, and Smith, and of McKenna's more specific conversational account, but which excludes any notion of desert, whether basic or non-basic. The point of blaming and praising on this notion largely forward-looking: its main objectives are protection, reconciliation, and moral formation. Agents are blameworthy and praiseworthy by virtue of being appropriate recipients of blame and

praise given these aims. Blaming on this conception can involve causing harm, but the justifiability of such harming does not reintroduce the legitimacy of desert. The resulting notion of moral responsibility is immune to any threat from the causal determination of action.

References

- Bok, H. 1998. *Freedom and Responsibility*, Princeton: Princeton University Press.
- Broad, C. D. 1952). "Determinism, Indeterminism, and Libertarianism," in his *Ethics and the History of Philosophy*, London: Routledge and Kegan Paul, pp. 195-217.
- Clarke, R. 2013. "Some Theses on Desert," *Philosophical Explorations* 16, pp. 153-64.
- Copp, D. 2008. "'Ought' Implies 'Can', and the Derivation of the Principle of Alternate Possibilities" *Analysis* 68, pp. 67-75.
- Dennett, D. 1984. *Elbow Room*, Cambridge: MIT Press.
- Dennett, D. 2003. *Freedom Evolves*. New York: Viking.
- Feinberg, J. 1970. *Doing and Deserving* . Princeton, NJ: Princeton University Press.
- Fischer, J. M., and M. Ravizza. 1998. *Responsibility and Control, A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- Ginet, C. 1966. "Might we have no Choice?" in *Freedom and Determinism*, Keith Lehrer, ed. New York: Random House.
- Ginet, C. 1990. *On Action*, Cambridge: Cambridge University Press.
- Haji, I. 1998. *Moral Appraisability*, New York: Oxford University Press.
- Haji, I. 2002. *Deontic Morality and Control*, Cambridge: Cambridge University Press.
- Haji, I. 2012. *Reason's Debt to Freedom*, New York: Oxford University Press.

- Hieronymi, P. 2001. "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, pp. 529-54.
- Harman, G. 1977. *The Nature of Morality*, Princeton: Princeton University Press.
- Hobbs, J. 2012. Dissertation, Department of Philosophy, Cornell University.
- Honderich, T. 1988. *A Theory of Determinism*, Oxford: Oxford University Press.
- Humberstone, L. "Two Sorts of 'Oughts'," *Analysis* 32, pp. 8-11.
- Hume, D. 1739/1978. *A Treatise of Human Nature*, Oxford: Oxford University Press.
- Kant, I. (1797/1963). *The Metaphysical Elements of Justice*, New York: Bobbs-Merrill, tr. John Ladd, pp. 99-107.
- Manne, K. (2011). *Not by Reasons Alone*, MIT, PhD Dissertation.
- Marcus R. B. 1966. "Iterated Deontic Modalities," *Mind* 75 580-82.
- McKenna, M. 2012. *Conversation and Responsibility*, New York: Oxford University Press.
- Nelkin, D. 2011. *Making Sense of Moral Responsibility*, Oxford: Oxford University Press.
- Nelkin, D. (forthcoming). "Comments on Michael McKenna's *Conversation and Responsibility*," *Philosophical Studies*.
- Norcross, A. 2006. "Reasons without Demands: Rethinking Rightness," in *Blackwell Contemporary Debates in Moral Theory*, J. Dreier, ed., Oxford: Blackwell.
- Pereboom, D. 1995. "Determinism *Al Dente*," *Noûs* 29, pp. 21-45
- Pereboom, D. 2001. *Living without Free Will*, Cambridge University Press.
- Pereboom, D. 2013a. "Free Will Skepticism, Blame, and Obligation," in *Blame: Its Nature and Norms*, Neal Tognazzini and D. Justin Coates, eds., New York: Oxford University Press, pp. 189-206.
- Pereboom, D. 2013b. "Moral Responsibility without Desert," for *Free Will and Moral*

- Responsibility*, Ishtiyaque Haji and Justin Caouette, eds., Newcastle upon Tyne: Cambridge Scholars Press, pp. 213-28.
- Pereboom, D. 2014. *Free Will, Agency, and Meaning in Life*, Oxford: Oxford University Press.
- Priestley, J. 1788. *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity, In a Correspondence between Dr. Price and Dr. Priestley*, Part III, pp. 147-152; reprinted in Joseph Priestley, *Priestley's Writings on Philosophy, Science, and Politics*, John Passmore, ed., New York: Collier.
- Rosch, E. 1972. "Universals in Color Naming and Memory," *Journal of Experimental Psychology* 93, p. 10-20.
- Rosch, E. 1973. "Natural Categories," *Cognitive Psychology* 4, pp. 328-50.
- Scanlon, T. M. 1998. *What We Owe to Each Other*, Cambridge MA: Harvard University Press.
- Scanlon, T. M. 2009. *Moral Dimensions*, Cambridge: Harvard University Press.
- Scanlon, T. M. 2013. "Giving Desert Its Due," *Philosophical Explorations* 16, pp. 101-16.
- Schlick, M. 1937. "When is a Man Responsible?" in *Problems of Ethics*, D. Rynin, tr., New York: Prentice-Hall, pp. 143-56.
- Schroeder, M. 2011. "Ought, Agents, and Actions." *Philosophical Review* 120, pp. 1-41.
- Sher, G. 2006. *In Praise of Blame*, Oxford: Oxford University Press.
- Smart, J. J. C. 1961. "Free Will, Praise, and Blame," *Mind* 70, pp. 291-306.
- Smith, A. 2008. "Control, Responsibility, and Moral Assessment," *Philosophical Studies* 138, pp. 367-392.
- Stapleton, Sean. 2010. *Hard Incompatibilist Challenges to Morality and Autonomy* Dissertation, Department of Philosophy, Cornell University, Ithaca NY.
- Strawson, G. 1986. *Freedom and Belief*, Oxford: Oxford University Press.

Strawson, G. 1994. "The Impossibility of Moral Responsibility," *Philosophical Studies* 75, pp. 5-24.

Strawson, P. F. 1962. "Freedom and Resentment," *Proceedings of the British Academy* 48, pp. 1-25.

van Inwagen, P. 1983. *An Essay on Free Will*, Oxford: Oxford University Press.

Vilhauer, B. 2013. "Persons, Punishment, and Free Will Skepticism," *Philosophical Studies* 162: 143-63.

Waller, B. 1990. *Freedom without Responsibility*, Philadelphia: Temple University Press.

Waller, B. 2011. *Against Moral Responsibility*, Cambridge MA: MIT Press.