Free Will Skepticism, Blame, and Obligation

Derk Pereboom, Cornell University

in *Blame: Its Nature and Norms*, Neal Tognazzini and D. Justin Coates, eds., New York:

Oxford University Press, pp. 189-206.

Penultimate draft

Abstract: I have argued we are not free in the sense required for moral responsibility in the

sense that involves basic desert, but that a conception of life without this type of free will

would not be devastating to morality or to our sense of meaning in life, and in certain respects

it may even be beneficial (Pereboom 1995, 2001, 2007). Here I set out a notion of blame and

its place in the moral life consistent with a denial of the sort of free will required for moral

responsibility in the basic desert sense, and I specify the respect in which this sort of blame

requires that its target have an ability to do otherwise.

Free will skepticism and blame

A central concern in the historical free will debate is whether the sort of free will

required for moral responsibility of a central kind is compatible with the causal determination

of our actions by factors beyond our control. Since Hume, this concern has prominently been

extended to whether this sort of free will is compatible with indeterminacy in action (Hume

1

1739/1978). On the skeptical view I endorse, free will, characterized in this way, is incompatible with this type of causal determination, but also with the kind of indeterminacy of action that Hume envisioned (Pereboom 1995, 2001, 2007). It is crucial, however, to recognize that the term 'moral responsibility' is used in a variety ways, and that the type of free will or control required for moral responsibility in some of these senses is uncontroversially compatible with the causal determination of action by factors beyond our control, which have therefore not been at issue in the debate between compatibilists and incompatibilists. It is potentially open to the free will skeptic to affirm that we are morally responsible in these other senses.

The particular variety of moral responsibility that has been at issue in the historical debate is set apart by the notion of *basic desert*. For an agent to be morally responsible for an action in this sense is for it be hers in such a way that she would deserve to be the recipient of an expression of moral resentment or indignation if she understood that it was morally wrong, and she would deserve be the recipient of an expression of gratitude or praise if she understood that it was morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve to be the recipient of the expression of such an attitude just because she has performed the action, given sensitivity to its moral status; not, for example, by virtue of consequentialist or contractualist considerations. One might be the recipient of the expression of such an attitude when this expression is either covert or overt. Moral responsibility in this sense is presupposed by our attitudes of resentment, indignation, gratitude, and moral praise, since having such an attitude essentially

2

involves the supposition that the agent in question basically deserves to be the recipient of its expression. It is thus the variety of moral responsibility that P. F. Strawson brings to the fore in his "Freedom and Resentment" (1962).

The free will skeptic will call into question any blaming practice that presupposes that the agent being blamed is morally responsible in the basic desert sense. Since much actual human blaming has this presupposition, any skeptical account of blame will be revisionary. At the same time there are a number of recent analyses of blame and blameworthiness that incompatibilists would not regard as conflicting with determinism, and that the free will skeptic can take on board (Pereboom 2009). For instance, in 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). The free will skeptic can, without any inconsistency, endorse such beliefs and desires about badness. One might object that if we gave up the belief that people are blameworthy in the basic desert sense we could no longer legitimately judge any actions as good or bad. But denying basic desert blameworthiness would not threaten axiological judgments of moral badness, and, likewise, denying basic desert praise- or creditworthiness would not jeopardize such assessments goodness. Even if we came to believe that a perpetrator of genocide was not blameworthy in this sense due to some degenerative brain disease, we could still legitimately maintain that it was extremely bad that he acted as he did. So far, then, the free will skeptic can accept blame on this analysis. Sher does maintain that blame involves a set of affective and behavioral dispositions, and at this point one might think

his account conflicts with free will skepticism. But he does not contend that any such dispositions are essential to blame, but rather only that they are connected to it in a weaker sense. Given the looseness of this tie, the free will skeptic can endorse blaming on Sher's specification. She might not endorse all of the affective dispositions that might be associated with blame – in particular, not those dispositions to expressions of moral resentment and indignation justified in virtue of basic desert. Still, two dispositions to which Sher draws our attention -- "to apologize for our own transgressions and vices and to reprimand others for theirs" (2006: 108), are compatible with the free will skeptic's conviction.[1]

In Thomas Scanlon's analysis, to blame an agent for an action is to judge that it reveals something about the agent's attitude toward oneself and/or others that impairs the relations that he can have with them, and to take one's relationship with him to be modified in a way that this judgment of impaired relations justifies as appropriate (Scanlon 2009: 128-31). Whether blame defined in this way is compatible with free will skepticism depends on how the appropriateness to which this characterization refers is understood. If it is taken to introduce basic desert, then the resulting notion of blame would be called into question. But there is an epistemic or evidential interpretation that is consistent with free will skepticism. One of Scanlon's examples illustrates this reading. You trusted Bill, but you then noticed that he repeatedly behaved in an untrustworthy manner, as a result of which it is now appropriate for

---

[1] This is not to say that I accept Sher's theory as a complete skeptical account of blame. For further critical review of Sher's account from other perspectives, see McKenna and Vadakin (2008), Reis-Dennis (2010), and Hobbs (2013).

you to take your relationship with him to be impaired to the degree that reflects this diminished trust. Here the justification is evidential. You believed Bill was trustworthy to a high degree, but you then acquired good evidence of reduced trustworthiness, and hence a good reason to judge that an attitude of his was relationship-impairing. You now make this judgment, and take your relationship with him to be impaired to a degree that it justifies as appropriate. All of this is unobjectionable to the free will skeptic. But let me note that blame in human relationships would seem only seldom linked with a judgment of relationship impairment. A fair proportion of blame is directed toward children by parents and teachers in the process of moral education, and only in a small minority of cases does such blame involve a judgment that a relationship has been impaired.[2]

Although Sher's and Scanlon's notions of blame can be accepted by the free will skeptic, I hope to secure a notion of blame that has a more pronounced forward-looking aspect. Free will skeptics like Joseph Priestley (1788), and their revisionary compatibilist cousins such as Moritz Schlick (1939) and J. J. C. Smart (1962), claim that given determinism, a kind of blame that can be retained is indeed forward-looking: the justification and goal of such determinism-friendly blame is to moderate or eliminate dispositions to misconduct. The dispositions addressed already exist and are manifested in past actions, and are reasonably assumed to persist unless corrective measures are taken. Such blaming, then, addresses past misconduct as

---

[2] This sort of criticism of Scanlon's proposal is developed by Wolf (2011) and Hobbs (2012); for further criticisms see Nelkin (2011a) and Wallace (2011),

a means to targeting such a standing disposition (Hieronymi 2001).[3] Thus 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 the point of asking such questions is to have him recognize and

acknowledge a disposition to behave immorally. If the reasons given in response to such

questions indicate that he does have such a disposition, it then becomes apt to request an

effort to eliminate it. 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 defended by Scanlon (1998) and Hilary Bok (1998).

Michael McKenna (2012) has recently developed a view that turns out, with some

revision, to be amenable to these goals; he calls it a *conversational* theory of moral

responsibility. In his conception, the actions of a morally responsible agent are potential

bearers of a species of meaning insofar as they are indicators of the quality of will with which

agents act (2012: 92-4). My blaming someone is an expression of a sentiment or attitude such

as resentment or indignation, and its function is to communicate my moral response to her in

the light of the indicated quality of will. When a morally responsible person acts, she

understands that members of the moral community might assign such a meaning to her action.

When her acts are morally charged, she appreciates that she might be introducing a

---

[3] Pamela Hieronymi (2001: 546) argues that resentment is best understood as a protest;

"resentment protests a past action that persists as a present threat." Although resentment is

not a feature of the forward-looking notion of blame I set out here, in this view a core function

of blaming someone is to protest a past action of his that persists as a present threat.

meaningful contribution to such a conversational exchange with others. This initial stage McKenna calls *Moral Contribution*. The second stage, in which that agent is blamed by a respondent, he calls *Moral Address*. In the third stage, *Moral Account*, the blamed agent extends the conversation by offering an excuse, a justification, or an apology, for example. The respondent might at this point continue the conversation perhaps by forgiving or punishing. In a further stage the blamed agent may be restored to full status in the moral community. This *moral responsibility exchange* is modeled on analogy with an ordinary conversational exchange between speakers of a natural language. Not all blaming fits this model; blaming the dead, for example, does not. Here McKenna invokes – plausibly, to my mind -- a paradigm-similarity model for the meaning and extension of the concept (Rosch 1972, 1973). The blame conversation is the central case of blame, and examples of other sorts count as blaming because they are sufficiently similar to the central case.

Some modification to this picture is in order.[4] When overt 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 in all cases in which an agent is overtly blamed, since frequently agents are blamed due to mistaken reports of having acted wrongly, or to negligent or

---

[4] McKenna is open to some of these amendments (in correspondence, cf. McKenna 2012: 90-91), but it is essential to his proposal that at the second stage the respondent adopts a blaming attitude, which might be defeated in the third stage. My view is that even though in practice people often do adopt a blaming attitude at the second stage, doing so typically involves moral error.

deliberate misinterpretations of actions, or to outright fabrication. Now McKenna envisions

that blame proper occurs at the second stage, while some theorists – Feinberg, for example –

propose that blame is a form of punishment, which in McKenna's model occurs at some stage

after the third. Overt blame might in fact occur at either juncture, that is, either prior to or

subsequent to a request for excuse, justification, or some other type of exonerating

explanation. But if it occurs prior to such a request, it is very often in moral error. If at the

second stage excuse or justification is significantly epistemically possible, an address that

communicates that the agent acted badly and without excuse or justification – that is, an

accusation -- will typically be morally erroneous. If an accusation is made at the second stage,

and a valid justification or excuse is offered, an apology on the part of the respondent is clearly

in order. But even if it was significantly epistemically possible that the agent had a valid excuse

or justification but in fact had none, a prior accusation would have been a moral mistake. In

the normative case, if it is evident at the second stage that the agent did perform the action in

question, the respondent will inquire whether he has an excuse or justification, and if it is not

obvious that the agent has performed the action, the respondent will attempt to determine

whether he did.

   If at the third stage the agent admits to having performed the action without excuse or

justification, then at the next stage the respondent might address the agent by saying to him

"What you did was wrong!" in a way that expresses moral resentment or indignation. This

would arguably be a form of blame that counts as punishment because it involves an intention

to inflict psychological pain in response to a perceived wrong done (cf., Wallace 1994: 51-83;

Nelkin 2011b: 31-50). If the agent does not admit to having performed the action, but it is nevertheless clear that he has done so, and without excuse or justification, the respondent might accuse him of having done wrong. This accusation would be a form of blame, and if it communicated in a way that expresses a reactive attitude of resentment or indignation, it would again arguably count as a form of punishment. But the respondent might also address the agent by expressing disappointment or sadness, without at the same time expressing resentment or indignation, and she might also counsel apology and reform, and this would not conflict with the claims of the free will skeptic.

As I said, the free will skeptic can endorse an answerability notion of moral responsibility, and she can accept an amended version of McKenna's more specific proposal insofar as it avoids an appeal to basic desert and to expressions of the desert-involving reactive attitudes. McKenna argues, convincingly, that his general model is compatible with both endorsing and rejecting basic desert. I propose, consistent with free will skepticism, to ground this model for blame not in basic desert, but in three non-desert invoking moral desiderata: protection of potential victims, reconciliation to relationships both personal and with the moral community more generally, and moral formation. Immoral actions are often harmful, and those who are potentially harmed have a right to protect themselves and to protect others from those who are disposed to behave harmfully. Such behavior can also impair relationships, and we have a moral interest undoing such impairment through reconciliation. And since we value morally good character and resulting action, we have a stake in the formation of moral character when it is plagued by dispositions to misconduct.

9

Blaming on such a conversational model can have a role in realizing each of these aims. Suppose someone acts badly, say by disseminating defamatory fabrications about his political rivals. One might then confront him by asking him what good evidence he has for his allegations, and supposing he cannot produce it, one might point out to him that his behavior is immoral and that he should cease to be disposed it. As McKenna points out, citing Feinberg, this process might harmful to the wrongdoer in various ways – it might cause psychological pain, for instance (McKenna 2012: 125; Feinberg 1970: 60-74; cf., Zimmerman 1988: 155). But inflicting this harm is justified first of all by the right of the defamed to protect themselves and to be protected from this type of aggression and its consequences. In addition, one might have a stake in reconciliation with the wrongdoer, and calling him to account can function as a step toward realizing this objective. Finally, we also have an interest in his moral formation, and the moral address described naturally functions as a stage in this process. Blame grounded in this way is essentially forward-looking, since its aims are future protection, future reconciliation, and future moral formation. The immediate object of blame is often a past action, but insofar as the purpose of blame is protection and moral formation, the past action will be addressed as a means for correcting a persisting disposition to act badly. Insofar as the goal of blame is reconciliation, the past action will also be addressed for its own sake.[5] There may be cases in which an immoral action has been performed but without a persisting disposition so to act, and blame can still have the point of reconciliation in such a case.

In accord with the conversational model, it is the agent's rationality that is engaged in

---

[5] Thanks to Dana Nelkin for making this point.

the envisioned process. At the stage of moral address, we ask for an explanation with the intent of having the agent acknowledge a disposition to act badly, and then, if he has in fact so acted without excuse or justification, we aim for him come to see that the disposition issuing in the action is best eliminated. In the standard sort of case, this change is produced by way of the agent's recognition of moral reasons to make it. Accordingly, it is an agent's responsiveness to reasons (cf. Fischer and Ravizza, 1998), together with the fact that we have a moral interest in our protection, his moral formation, and our reconciliation with him, that explains why he is an appropriate recipient of blame in this forward-looking sense.

McKenna contends that there is a close tie between blame on the conversational model and basic desert. In his view, the 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. (2012: 150)

The harms upon being blamed include the emotional pain of engagement in the moral conversation, and the alteration for the worse of interpersonal relationships. McKenna proposes three respects in which such harm might be understood as a noninstrumental good, and as a result, basically deserved. For each, my sense is that the good on offer turns out to be only instrumental to realizing some further good, and thus not basically deserved. The first is the good "located in the blameworthy agent's commitment to membership within the moral

11

community." On the basic desert view, it is a good that the blameworthy agent is harmed in the ways indicated just because he has knowingly done wrong, and in the context of the debate, this is what it is for such a harm to be a noninstrumental good. Harm aimed at the good of membership in a moral community would, by contrast, be instrumental, since the harm is not envisaged as a good in itself, but instead as serving the good of such membership. The second respect draws on Sher's view that "when one blames someone, her desire that a person not have failed to live up to morality's counsel is explained by her caring for these aspects of morality quite generally." McKenna contends: "we can say that conversing with a wrongdoer by blaming is motivated out of a commitment to morality. This is to be valued; it is a good" (2012: 168). But here again the harm is instrumental, since it is not conceived as a good in itself, but rather as serving the commitment to morality. Even if the good of morality is a noninstrumental good, the harm motivated by a commitment to morality will be an instrumental good, since it will be instrumental relative to the commitment to morality. The third respect McKenna cites is that "the goodness of the activity of blaming… concerns the noninstrumental value of a process that begins at one end with a wrong done, that then conversationally answers that wrong by way of some blaming practice, and that invites an extension of the unfolding conversation in a manner that values sustained bonds of moral community" (2012: 169). There are two ways that this explanation can be construed as adducing harm as an instrumental good. One is that the claimed noninstrumental good at play is not the harm, but rather the conversational process; the harm should then be understood as a good instrumental to the larger good of the overall process. The other is that harm serves the

noninstrumental good of the sustained bonds of the moral community, but then again the harm turns out not to be a good in itself, but rather as serving the good of these bonds. The free will skeptic can endorse each of these ways in which harm incurred by overt blame is a good instrumental relative to a further good; each can be understood as a way in which blame can serve the good of moral formation.

McKenna replies by contending that the harm of blame is partially constitutive of the various noninstrumental goods he adduces, and thus not merely instrumentally valuable (in correspondence; cf. McKenna 2012: 123-24, n. 18). But while it is plausible that certain kinds of obvious goods, like 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 a prerequisite of physical health, and health constitutive of flourishing, but it's not plausible that the pain of vaccination is constitutive of flourishing, by contrast with being instrumentally required for it. So my sense is that the proposed connection between moral responsibility on a conversational model and harm as a noninstrumental good hasn't been substantiated, and that thus no close tie between responsibility on this model and harm as basically deserved has been established.

Blame and obligation

Against this skeptical account one might object that for an agent to be blameworthy even in this forward-looking sense requires that she ought not to have performed the action,

and this in turn requires that she could have avoided the action, which is incompatible with determinism. While my free will skepticism does not endorse determinism, it leaves open determinism as a serious possibility. And according to the consequence argument determinism is incompatible with being able to do otherwise, at least in one salient sense, and on this understanding I find the argument attractive (van Inwagen 1983; Ginet 1990: 95-117). So this objection poses a threat to my view.

Here is Dana Nelkin's version of the argument that fuels the objection:

N1. If S is blameworthy for A, then S ought not to have done A.

N2. If S ought not to have done A, then S could have refrained from doing A. (from OIC)

C. So, if S is blameworthy for A, then S could have refrained from doing A. (Nelkin 2011b: 100-101; cf. Copp 2008).

In assessing this argument one must tread with caution, since 'ought' has a range of correct uses, and it may be that not all are linked to the same OIC requirement, and that some are not associated with such a requirement at all. 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.[6] An 'ought to do,' Harman proposes, "implies that an

---

[6] Cf., Kate Manne (2011). In Mark Schroeder's characterization, the *deliberative* sense of 'ought' expresses a relation between an agent and an action, where an action is a property of an agent, while the *evaluative* 'ought', as in 'Larry ought to win the lottery' where Larry has been subject to a series of undeserved misfortunes. The evaluative 'ought' expresses relation to a proposition, in this case *that Larry wins the lottery* (Schroeder 2011). Manne argues that it is important to see

agent has a reason to perform an action, whereas an 'ought 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). But as

James Hobbs (2012) points out, certain 'ought to be' claims will have implications for reasons

agents have to act. He suggests the following diagnosis: 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. He contends, plausibly to my mind, that this assumption about reasons for action does

not imply a route, accessible to the agent, to the realization of what ought to be. But if, by

contrast, an agent *ought to do* something, that agent has a reason to do it, and there is such a

route to what she ought to do. In a similar vein, Nelkin (2011b) argues that 'ought' propositions

that specify what an agent ought to do are essentially action-directed, so that if 'S ought not to

do A' is true, then as a matter of the meaning of 'ought' propositions, or of the essential nature of

obligation, S is thereby directed to refraining from A, and this entails that S can refrain from A.

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 (at least directly) entail a 'can'

claim, while an 'ought to do' expresses a demand of an agent in a particular circumstance,

which does entail that the agent can perform the indicated action (cf., Humberstone 1971;

Manne 2011). One might call this second type an 'ought' of *specific agent demand*.

Hobbs's more precise proposal for the ability entailed by an action-directed 'ought' is

---

that the evaluative 'ought' applies not only to non-agential states of affairs, but also to actions.

that if a person ought all-things-considered to do something, then she has the physical and mental ability, the skill, and the know-how needed, and she is in circumstances appropriate for doing that thing. However, she may not have the motivation required to do it.[7] 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 determinism 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, not with the remaining crucial and causally relevant feature, her motivation. Hence, 'S ought not to do A' can be true in a deterministic situation in which S performs immoral act A because he is not appropriately motivated.

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 possibly be motivated. He responds by arguing that it is not exactly clear what this stronger sense of 'ought' is supposed to 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

---

[7] In Hobbs's account (2012), S has a basic ability to do A if and only if S is intrinsically such that if S were sufficiently motivated to A and S had an opportunity to A, then S would be relatively likely to A (taking into account the difficulty of A-ing).

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 it's my sense that if causal determinism precludes alternative possibilities for motivation and hence for action, there is a core sense of 'ought to do' that will be compromised. If I know that at some time an agent could not have avoided lacking the motivation required for performing some morally exemplary action, it would be unfair, and I think, mistaken, for me to claim that she ought to have performed that action at that time.

Still, it might well not be unfair for me to recommend to that agent that she perform an action of that type at some future time, given that it is epistemically open that she will develop the requisite motivation by then, and in particular if the recommendation might plausibly contribute causally to producing the motivation. And 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. The sense of 'ought' invoked here is thus plausibly distinct from the 'ought' of specific agent demand. I propose that, given determinism and that determinism precludes alternatives, when one tells an agent that she ought to perform an action of some type in the future, it is not the 'ought' of specific agent demand, but rather an 'ought' of axiological evaluation or ideality that would need to be invoked, but in a way that uses it to recommend the action to the agent. Such a use of 'ought' proposes to an agent as morally valuable a state of affairs in which she performs an action of a certain sort and recommends that she perform this action. We might call this the 'ought' of *axiological recommendation*. Given determinism and that determinism rules out alternatives, the use of

the 'ought' of specific agent 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

be typically untrue, while the 'ought' of axiological recommendation would not be similarly

undermined. Like the 'ought' of specific agent demand, the 'ought' of axiological

recommendation necessarily 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, via reasons for action, to her acting in some relevant way. One

might be unsure about whether such a route is accessible, while the use of 'ought' is

nevertheless legitimate.

So supposing that this world is deterministic, and given that determinism does preclude

the relevant alternative possibilities, and we typically don't know how agents will be

motivated, a correct and fair use of the prospective 'ought' of specific agent demand would be

compromised. Yet the 'ought' of axiological recommendation would often be in the clear.

Further, this use of 'ought' meshes nicely with the forward-looking notion of blame grounded

in the goods of protection, reconciliation, and moral formation. If the 'ought' in

N2. If S ought not to have done A, then S could have refrained from doing A

is read as the 'ought' of specific agent demand, then I, like Nelkin, accept N2. But this does not

rule out the legitimate prospective use of 'ought' claims of axiological recommendation in cases

where it turns out that due to determinism the agent could not have performed the act specified

at the time she failed to do so, in particular if there was reason to believe that the agent could

develop the requisite motivation, and especially if the articulation of the 'ought' claim could

reasonably have been expected to contribute to producing it.[8]

How do these reflections bear on the relation between blameworthiness and 'could have done otherwise'? Here again is Nelkin's argument:

N1. If S is blameworthy for A, then S ought not to have done A.

N2. If S ought not to have done A, then S could have refrained from doing A. (from OIC)

C. So, if S is blameworthy for A, then S could have refrained from doing A.

First, since I defend Frankfurt examples against an alternative possibilities requirement for moral responsibility, I reject the conclusion on one key reading (Pereboom 2001, 2012).[9] Specifically, I deny that in the basic desert sense of blameworthiness, if S is blameworthy for A, then S's

---

[8] That said, it might at times be misleading to use such 'ought to do' formulations under these conditions, since they are reasonably interpreted as specific agent demands. Formulations that are more clearly expressions of axiological recommendation, such as 'it would be good of you to do A', would preclude misinterpretation. Still, as Neal Tognazzini and Justin Coates point out (in correspondence), in "You know, you really ought to do A," 'ought to do' would seem to have axiological import.

[9] In Frankfurt examples (Frankfurt 1969), an agent considers performing some action, but an intervener, say a neuroscientist, is concerned that she will not come through. So if she were to manifest an indication that she will not or might not perform the action, the neuroscientist would intervene. But as things actually go, the neuroscientist remains idle, since the agent performs the action on her own. The idea is that even though the agent could not have avoided the action she performs, she is still intuitively morally responsible for this action.

blameworthiness will be explained per se by the fact that she could have refrained from doing A. But I accept the second premise, N2, on the specific agent demand reading of 'ought.' As a result, I'm forced to reject the first premise, N1, given the basic desert sense of blameworthiness and the specific agent demand reading of 'ought.'

How might this be explained? When an agent is tempted to perform immoral action A, one might say to him, using the 'ought' of specific agent demand: "You ought to avoid doing A!" Suppose he does A anyway. If the source incompatibilist defender of Frankfurt is right, and I argue that he is (Pereboom 2001), then an agent can be blameworthy in the basic desert sense if he is the appropriate kind of indeterministic source of his action even if he could not have avoided the action. The 'ought' judgment, however, directs him to the avoidance of A, and in my view to do so would involve a mistake and would be unfair if he could not have avoided A. So for the source incompatibilist Frankfurt-defender, the general claim

N1. If S is blameworthy for A, then S ought not to have done A

will be false, supposing blameworthiness in the basic desert sense, and the 'ought' of specific agent demand. Following Ishtiyaque Haji (1998), the advocate of this position can endorse

N1*. If S is blameworthy for A, then for S doing A is bad,

as an alternative account of the relation between basic desert blameworthiness and immoral action. N1* specifies an axiological prerequisite for blameworthiness, and not, as in N1, a deontological constraint. That it is bad for S to have done A does not require that S could have done otherwise.

But the source incompatibilist free will skeptic can endorse N1, and the soundness of the

subsequent argument, if it is read as invoking the 'ought' of axiological recommendation together with blameworthiness in our forward-looking sense:

> N1'. If S is blameworthy for A in the forward-looking sense, then it is appropriate for a relevantly positioned respondent T to tell S that he ought, in the sense of axiological recommendation, not perform actions of A's type.
>
> N2'. If it is appropriate for T to tell S that he ought, in the sense of axiological recommendation, not to perform actions of A's type, then it is open for T that S could, in the future, refrain from performing actions of A's type.
>
> C'. So, if S is blameworthy for A in the forward-looking sense, then it is open for T that S could, in the future, refrain from performing actions of A's type.[10]

I accept the argument set out in this way. Blameworthiness in the forward-looking sense licenses the right sort of respondent to tell the agent who has acted badly that he ought not to act this

---

[10] One might ask whether on the resulting view there are any genuine moral requirements for agents. Perhaps the answer is negative, since legitimate 'ought' judgments will be turn out to be axiological recommendations. A skeptic might then deny we are ever obligated, in the sense of specific agent demands, to do anything, and instead defend a purely axiological ethics. Alastair Norcross (2006) proposes such a view, one that rejects any deontology, even the maximization demand of the classical utilitarian position. Instead, accessible states of affairs are ranked in accord with value, and in moral practice, agents are encouraged to act with the aim of realizing a state of affairs reasonably high on the list.

way, where 'ought' has the sense of axiological recommendation, which in turn requires that it be epistemically open for the respondent that the agent can comply with this recommendation. While it's plausible that an agent's blameworthiness in the basic desert sense for an action is not explained per se by the fact that he could have avoided it, for an agent to be blameworthy for an action in the forward-looking sense it must be open that he refrain from performing actions of this type in the future. The Frankfurt-defender can accept this conclusion. It's intuitive that the agent in a typical Frankfurt example is blameworthy in the basic desert sense, but also in the forward-looking sense. But such an agent will then retain a general ability to refrain from performing the action at issue (Smith 2003; Vihvelin 2004; Fara 2008; Nelkin 2011b: 66-76) even though due to the intervention set-up he cannot exercise that ability at the time (Clarke 2009). This leaves it open that he will exercise that ability in the future. When a Frankfurt case is constructed so as to permanently preclude the exercise of this ability (Pereboom 2001: 27-8), then it is evident that the agent is not blameworthy for performing the action in the forward-looking sense.

Blame without the reactive attitudes

What form will blame take absent the basic desert presupposing reactive attitudes, such as moral resentment and indignation, and will it be effective? It is often supposed that blaming behavior would be morally optimal only if it involves these reactive attitudes. I argue, first, that there are alternative attitudes that do not presuppose basic desert that can be as effective morally as these reactive attitudes, and second, that in certain important respects

blame without these reactive attitudes is to be preferred (Pereboom 2001, 2009; cf. Honderich 1988).

Let us combine the negative basic desert involving reactive attitudes under the heading 'moral anger.' One might note that moral anger plays an important communicative role in our personal and societal relationships, and object that if we were to strive to modify or eliminate this attitude, such relationships might well be worse off. But when someone is mistreated in a relationship there are other emotions typically present that are not threatened by the skeptical view, whose expression can also communicate the relevant information. These emotions include feeling disappointed, hurt or shocked about what the offender has done, moral concern for him, and moral sadness and sorrow generated by this concern when the harm done is serious (Pereboom 2001, 2009). Parents might feel intensely sad, and not angry, that their son has driven his car while intoxicated and injured a pedestrian, and be concerned about his moral commitment. Ordinary human experience indicates that communicating such sadness and concern can be an effective way to motivate avoidance of future misbehavior. Often communicating anger is not required in addition to secure this effect. Feigned moral sadness is sometimes used to manipulate others, but what I have in mind is the genuine version. The alternatives to moral anger are not aggressive in the way that anger can be, and all by themselves they do not typically have anger's intimidating effect. If aggressiveness or intimidation is required, a strongly worded threat, for instance, might be appropriate. It is thus not clear that moral anger is required, or optimal, for communication in interpersonal relationships.

Shaun Nichols (2007) contends that sadness is an inadequate substitute for moral anger in personal and social relationships. His justification begins with the claim that moral anger can be shown, by way of empirical studies, to be beneficial to human relationships in certain crucial respects. He then argues, also on the basis of empirical work, that sadness will be much less effective in achieving the benefits. The essential elements of my response are, first, that Nichols's argument is remiss in not counting the cost of moral anger in comparison with the proposed substitutes; and second, that the studies he cites fail to show that adult human beings, with education and resolve, would not benefit overall from these substitutions in their personal and social relationships.

First, Nichols argues that while there is ample evidence that moral anger discourages cheating, defecting, and mistreatment, sadness tends to produce no behavior at all. Here he cites Richard Lazarus, who claims: "In sadness there seems to be no clear action tendency—except inaction, or withdrawal into oneself" (Lazarus 1991: 251). The evidence Lazarus cites is from infancy research, according to which "infants show individual differences in their propensities to feel sad or angry when blocked from attaining a desired end—some babies are more likely to feel sad, others to feel angry." Nichols also argues that "researchers have found that when infants show sadness as their predominant emotion, this is associated with giving up (Lewis & Ramsey 2005, 518), and it seems to be akin to learned helplessness (Abramson et al. 1978)," and that "by contrast, infants who respond with anger are more likely to try to overcome the obstacle (Lewis & Ramsey 2005, 518)."

In response, the point I want to make about the appropriateness of sorrow and sadness as a substitute for moral anger concerns adults, and not infants. It is not surprising that infants would not have developed the capacities to have thoughts like "I'm sad about what my brother has done to me, and now I will try, diplomatically, to improve this relationship." Thoughts of this sort are available to adults. Consider cases of adult sadness, in the absence of anger, about states of affairs that could not have been prevented, such as a hurricane devastating one's town, or a severe illness of a child. It's clear that adults can take action, and typically do, under such circumstances by way of these kinds of motivating factors. Similarly, when a parent is too intensely sad to be angry about her son's intoxicated driving and hitting a pedestrian, experience indicates that this attitude can motivate her take measures to cause him change his behavior. My claim is that for adults, moral sadness and sorrow -- accompanied by a resolve for fairness and justice, or to improving personal relationships -- will serve societal and personal relationships as well as moral anger does.

A further point, due to Carla Bagnoli, is that a distinction can be drawn between sadness as a form of moral address and sadness that does not have this role.[11] When infants are sad about some state of affairs, it's dubious that their sadness could be intended to have the role of addressing another agent for his bad behavior and immoral dispositions. Infants have not reached the level of cognitive development to be capable of such moral address. Rather, it's like the sadness that we might feel when we've had to cancel the picnic due to rain, or we have to miss the important game due illness. When one's sadness is a form of moral

---

[11] In correspondence.

address, it is directed toward an agent in virtue of his misbehavior, and this sort of sadness, I would venture, can clearly be effective in eliciting reform and reconciliation.

Nichols (2007) considers my proposal to supplement of sadness with resolve, but here he argues that it is unlikely that resolve will provide sufficient motivation for most people. "After all," he says, "many teenagers think that they risk going to hell if they have sex, yet this often provides insufficient motivation for abstinence." He also cites the Marxist thought that "working hard will generate benefits for the state which will in turn benefit everyone," which he thinks "turns out to be naively optimistic about the plasticity of human motivation." He places the proposal to replace the reactive attitudes with sadness and resolve in the same camp. In response, Nichols may be right that desires for sex and personal material incentives for work are not evenly matched by resolve together with alternative motivations or attachment to abstract principles; (notice that in his teenage sex example, not even the threat of the most severe punishment imaginable proves especially effective). But we have reason to believe that we can effectively oppose behavior that hinders good personal and social relationships with a resolve to make the world more fair and just, or to improve one's personal relationships, together with attitudes other than moral anger, and measures other than punishment driven by such anger. Mahatma Gandhi and Martin Luther King are often cited as exceptionally successful at resisting injustice without expression of moral anger (Nelkin 2011b). In many parts of the world, children are raised and taught with much less expression of moral anger than they were a century ago and earlier. Life in primary schools in the nineteenth century, as described by Charles Dickens, for example, featured a far greater incidence of

expression of moral anger than it does today, with results that are at very least no worse. One should also note that for many centuries now human beings have developed communities in which training and teaching methods are employed to diminish moral anger, and to develop moral and religious excellence by other means; Buddhist and Christian monastic societies, and radical reformation groups such as the Amish come to mind. It's especially important that we examine such communities to see whether these alternative methods can be successful.

My second contention is that in certain respects blame without moral anger is preferable.

I have argued that destructive anger in relationships is nourished by the belief that its target is in this sense blameworthy for having done wrong. The anger that fuels ethnic conflicts often results partly from the belief that an opposing group so deserves blame for some atrocity. Free will skepticism advocates retracting such beliefs because they are false, as a result of which the associated anger might be diminished, and its expressions reduced (Pereboom 2001). In addition, expression of moral anger is more likely to occasion destructive resistance than is expression of moral concern and the moral sorrow that such concern is apt to generate. Sacha Sullivan (in conversation) points out that trainers of powerful animals are especially wary of expressing anger, since the reaction is so dangerous and less effective than available alternatives. One might reasonably believe that the analogous claim is true for relations among humans (Pereboom 2009). (Note that trainers' ability to avoid expressions of anger counts against the concern that we lack the general ability to avoid such anger in human relationships.)

But in addition, blame fueled by moral anger arguably renders it particularly susceptible to errors that threaten to undermine the integrity and effectiveness of the moral conversation. We see that parents who become angry with children when their actions produce bad consequences are disposed to believe that these consequences were intentionally caused even if it is open that they are merely accidental, that when we are angry with politicians whose views we find despicable we too readily believe reports that they've behaved immorally, and that human beings too easily come to believe accusations against members of groups with which they are angry for perceived past immoral behavior. In the last several decades, impressive experimental evidence that blaming behavior is widely subject to problems of these kinds has been mounting (Nadelhoffer 2006). Surveys conducted by Mark Alicke and his associates indicate that subjects who spontaneously evaluate agents' behavior unfavorably are apt to exaggerate their causal control and any evidence that might favor it while deemphasizing counterevidence (Alicke, Davis, and Pezzo 1994; Alicke 2000; Alicke, Rose, and Bloom 2012). Alicke calls this tendency 'blame validation.' Studies designed by Joshua Knobe (2003) and by Thomas Nadelhoffer (2004a, 2004b) indicate that subjects are much more likely to judge that a bad side effect was produced intentionally than they are to judge that a good side effect was produced intentionally in an otherwise structurally similar case. A survey conducted by Nadelhoffer provides evidence that initial judgments concerning the blameworthiness and praiseworthiness of an agent have similar influences on ascriptions of intentionality (Nadelhoffer 2004b).

One might question whether moral anger has a significant role in producing the sorts of

fabrications indicated by these surveys. Studies might be designed to determine whether this is

the case. My hypothesis is that this is indeed often so, while other impulses, such as a desire to

disempower or dominate, or sheer glee at another being blamed and punished, might well also

result in these kinds of effects. Still, my sense is that it is often moral anger has the key role.

Evidence for this includes the fact that when someone blames in a way that that expresses

anger we are typically on guard that he will attribute to its target intentions and efficacy

skewed in a way that would serve to justify the blame. It stands to reason that such effects

would be diminished if such anger were replaced by genuine moral concern for the agent and

the moral sorrow such concern would generate if the offense was especially serious. It is

noteworthy that in the various surveys in which such effects are indicated, a substantial

percentage of those surveyed does not display the sort of fabrication targeted. It would be

interesting to determine whether members of this group are less subject to moral anger, or are

capable of overcoming it by way of some rational or affective technique.

Final words

To this account one might object that unless a response to bad behavior expresses the

reactive attitudes of resentment or indignation it is not genuine blame (Wallace 2011). In

response, when a parent points out to a child that what he did was immoral, and recommends

that he not perform similar actions in the future, but does so without indignation but only

disappointment, it seems clear that no linguistic error is made when we say that under such

conditions the child is being blamed. Let me note again that probably most of the blame that

occurs in this world is blame of children by parents and teachers, and that in many contemporary contexts such blame does not feature moral anger. Most parents are morally angry only in cases in which the misbehavior is especially serious.

Moreover, consonant with the phenomenon of blame validation and its connection with the reactive attitudes, we seem often to make errors about intention and causation when moral anger is involved in blaming. Certain kinds of children's stories sometimes feature parents whose expression of moral anger is never or infrequently accompanied by such error, but my sense is that this is the stuff of fiction. I suspect that children have a keen awareness of this problem, and as a result are likely to discount the reasonableness of overt blame when it expresses reactive attitudes. They might be intimidated by anger-expressing blame just as they would be by a threat of physical violence, but this departs from the model of a rational moral conversation. As a result, we are apt to think that parents who express indignation whenever they believe that their children have misbehaved even in relatively minor ways fall short of the ideal in a significant respect.

Still, some types and certain degrees of moral anger are likely to be beyond our power to affect, and thus even supposing that the free will skeptic is committed to doing what is right and rational, she would still be unable to eliminate these attitudes. Nichols cites the distinction between narrow-profile emotional responses, which are local or immediate emotional reactions to situations, and wide-profile responses, which are not immediate and can involve rational reflection (Nichols 2007). As free will skeptics we should expect that we cannot avoid some degree of local and immediate moral anger. But in the long term, we might well be able

30

to take measures that would moderate or eliminate moral anger, and given a belief in free will skepticism, we would then do so for the sake of morality and rationality.[12]

References

Abramson, L., M. Seligman, and J. Teasdale. 1978. "Learned Helplessness in Humans: Critique and Reformulation," *Journal of Abnormal Psychology 87*, pp. 49–74.

Alicke, M. D. 2000. "Culpable Control and the Psychology of Blame, *Psychology Bulletin* 126, pp. 556-74.

Alicke, M. D., T. L. Davis, M. V. Pezzo. 1994. "A Posteriori Adjustment of A Priori Decision Criteria," *Social Cognition* 8, pp. 286-305.

Alicke, M. D., D. Rose, and D. Bloom. 2012. "Causation, Norm Violation and Culpable Control," *Journal of Philosophy*, forthcoming.

Bok, H. 1998. *Freedom and Responsibility*, Princeton: Princeton University Press.

Clarke, R. 2009. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind* 118, pp. 323-51.

David Copp, 2008. "'Ought' Implies 'Can', and the Derivation of the Principle of Alternate Possibilities" *Analysis* 68, pp. 67-75.

Fara, M. 2008. "Masked Abilities and Compatibilism, *Mind* 117, pp. 843-65.

Feinberg, Joel. 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.

Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, pp. 829-39.

Ginet, C. 1990. *On Action*, Cambridge: Cambridge University Press.

Haji, I. 1998. *Moral Appraisability*, 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.

Knobe, J. 2003. "Intentional Action in Folk Psychology: An Experimental Investigation,"

    Philosophical Psychology 16 (2), pp. 309-24.

Lazarus, R. 1991. *Emotion and Adaptation*, Oxford: Oxford University Press.

Lewis, M. and D. Ramsey. 2005. Infant Emotional and Cortisol Responses to Goal Blockage.

    Child Development 76, pp. 518–30.

Manne, K. (2011). *Not by Reasons Alone*, MIT, PhD Dissertation.

Marcus R. B. 1966. "Iterated Deontic Modalities," *Mind* 75 580-82.

McKenna M., and Vadakin, A. 2008. "Review of George Sher, *In Praise of Blame*," *Ethics* 118

    (4):751-756.

McKenna, M. 2012. *Conversation and Responsibility*, New York: Oxford University Press.

Nadelhoffer, T. 2004a. "The Butler Problem Revisited," *Analysis* 64, pp. 277-88.

Nadelhoffer, T. 2004b. "Praise, Side Effects, and Intentional Action," *Journal of Theoretical and*

Philosophical Psychology 24, pp. 196-213.

Nadelhoffer, T. 2006. "Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems

for Jury Impartiality," *Philosophical Explorations* 9, pp. 203-20.

Nelkin, D. 2011a. "Review of Thomas Scanlon, *Moral Dimensions*," *Philosophical Review* 120,

pp. 603-607.

Nelkin, D. 2011b. *Making Sense of Moral Responsibility*, Oxford: Oxford University Press.

Nichols, S. 2007. "After Compatibilism: a Naturalistic Defense of the Reactive Attitudes,"

Philosophical Perspectives 21, pp. 405-28.

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. 2007. "Hard Incompatibilism" and "Response to Fischer, Kane, and Vargas," In J.

Fischer, R. Kane, D. Pereboom, M. Vargas, *Four Views on Free Will*, Oxford: Blackwell

Publishers.

Pereboom, D. 2009. "Free Will, Love, and Anger," *Ideas y Valores: evista de Colombiana de*

Filosofía 141, pp. 5-25.

Pereboom, D. 2012. "Frankfurt Examples, Derivative Responsibility, and the Timing Objection,"

Philosophical Issues, forthcoming.

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.

Reis-Dennis, Samuel. 2010. *Blame*. Undergraduate Honors Thesis, Department of Philosophy,

Cornell University.

Russell, P. 2000. "Compatibilist-Fatalism", in *Moral Responsibility and Ontology,* T. van den

Beld, ed., Dordrecht: Kluwer, pp. 199-218.

Rosch. E. 1972. "Universals in Color Naming and Memory." *Journal of Experimental Psychology*

93, pp. 10-20.

Rosch, E. 1973. "Natural Categories," *Cognitive Psychology* 4, pp. 328-50.

Scanlon, T. 1998. What We Owe to Each Other, Cambridge MA: Harvard University Press.

Scanlon, T. 2009. *Moral Dimensions*, Cambridge MA: Harvard University Press.

Schlick, M. 1939. "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.

Smilansky, S. 2000. *Free Will and Illusion*, Oxford: Oxford University Press.

Smith, M. 2003. "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and

Compulsion, Sarah Stroud and Christine Tippolet, eds., *Weakness of Will and Practical*

Irrationality, Oxford: Oxford University Press.

Spinoza, B. 1677/1985. *Ethics*, in *The Collected Works of Spinoza*, E. Curley ed. and tr., Volume

1, Princeton: Princeton University Press.

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.

Vihvelin, K. 2004. "Free Will Demystified: A Dispositional Account," *Philosophical Topics* 32, pp. 427-50.

Wallace, R. J. 1994. *Responsibility and the Moral Sentiments*, Cambridge MA, Harvard University Press.

Wallace, R. J. 2011. "Dispassionate Opprobrium: On Blame and the Reactive Sentiments," in R. J. Wallace, R. Kumar, and S. Freeman, eds., *Reasons and Recognition. Essays on the Philosophy of T. M. Scanlon*, New York: Oxford University Press, pp. 348–72.

Wolf, S. 2011. "Blame, Italian Style", in R. J. Wallace, R. Kumar, and S. Freeman, eds., *Reasons and Recognition. Essays on the Philosophy of T. M. Scanlon*, New York: Oxford University Press, pp. 332-348.

Zimmerman, M. J. 1988. *An Essay on Moral Responsibility* . Totowa, NJ: Rowman and Littlefield.