Free Will, Love, and Anger

Derk Pereboom, Cornell University

Ideas y Valores: Revista de Colombiana de Filosofía 141, 2009, pp. 5-25.

Penultimate Draft

Abstract: I have argued we are not free in the sense required for moral responsibility, while at the same time 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 2001). In this article, I explore which sorts of emotional attitudes are consistent with a denial of the sort of free will required for moral responsibility, and whether they can sustain a meaningful life. In the process, I respond to Shaun Nichols's recent criticism of my position (Nichols 2007).

1. Moral responsibility and the reactive attitudes.[1]

A central concern in the historical free will debate is whether the sort of free will required for moral responsibility 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. On the skeptical view

[1] Thanks to Mike Patterson, Sacha Sullivan, Dana Nelkin, and Shaun Nichols for valuable commentary and discussion. This paper is a partial response to the challenge that Tamler Sommers sets out in his recent article, "More Work for Hard Incompatibilism" (Sommers 2009).

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. Here it is crucial 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 several of these senses is uncontroversially compatible with the causal determination of action by factors beyond our control. At the same time there is one particular sense of moral responsibility that has been at issue in the historical debate. It is this: for an agent to be morally responsible for an action is for it be hers in such a way that she would deserve blame if she understood that it was morally wrong, and she would deserve credit or perhaps 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 the blame or credit just because she has performed the action, given sensitivity to its moral status, and not by virtue of consequentialist or contractualist considerations. This characterization allows that agent be morally responsible for an action even if she does not deserve blame, credit, or praise for it -- if, for example, the action is morally indifferent. Moral responsibility in this sense is presupposed by our retributive reactive attitudes (Pereboom 2001) , and it is thus the variety of moral responsibility that P. F. Strawson famously brings to the fore in his "Freedom and Resentment" (1962). The type of moral responsibility that incompatibilists claim not to be compatible with determinism is the sense characterized by basic desert and the reactive attitudes that presuppose it. From this point on, unless otherwise indicated, I will use the term 'moral responsibility' to refer to this particular variety.

Let me characterize my position more precisely. Spinoza (1677/1985: 440-4, 483-4, 496-7) held that due to very general facts about the nature of the universe, we human beings lack the sort of free will required for moral responsibility. About this I think he is right. More specifically, he contends that it is because of the truth of causal determinism that we lack this sort of free will; he is thus a hard determinist. By contrast, I am agnostic about the truth of causal determinism. I argue, in agreement with Spinoza, that we would not be morally responsible if determinism were true, but also that we would lack moral responsibility if indeterminism were true and the causes of our actions were exclusively states or events – this is the notion of indeterminacy of action that Hume arguably had in mind (1739/1978). For such indeterministic causal histories of actions would be as threatening to this sort of free will as deterministic histories are. Still, it might be that if we were undetermined agent-causes – if we as substances had the power to cause intentions and decisions without being causally determined to cause them – we would then have this type of free will. But although our being undetermined agent causes has not been ruled out as a coherent possibility, it is not credible given our best physical theories. Thus I do not claim that our having the sort of free will required for moral responsibility is impossible. Rather, I don't take a stand on whether it is possible. Nevertheless, because the only account in which we are likely to have this kind of free will is not credible given our best physical theories, it is likely that we are not free in the sense required for moral responsibility. I call this position hard incompatibilism. However – and this is the issue I will address here -- I contend 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 (for opposing views, see Smilansky 2000; Russell 2000). Strawson denies this. In what follows, I will explore which sorts of emotional attitudes are consistent with a denial of the sort of free will required for moral responsibility, and whether they can sustain a meaningful life.

2. Moral Anger.

Strawson maintains that the justification for claims of blameworthiness and praiseworthiness is ultimately grounded in the system of human reactive attitudes, and since moral responsibility has this type of basis, the truth or falsity of causal determinism is not relevant to whether we legitimately hold agents morally responsible. Moral responsibility is founded in the reactive attitudes required for the kinds of relationships that make our lives meaningful. If causal determinism did threaten the reactive attitudes, we would face the prospect of a certain "objectivity of attitude," a stance that, in Strawson's conception, undermines the possibility of good interpersonal relationships (Strawson 1962). I think that Strawson is right to believe that objectivity of attitude would seriously hinder interpersonal relationships (for a contrasting view, see Sommers 2007), but that he is mistaken to hold that it would result or be appropriate if determinism did pose a genuine threat to the reactive attitudes.

First, some of our reactive attitudes, although they would be undermined by hard determinism, or more broadly by hard incompatibilism, are not required for good interpersonal relationships. Indignation and moral resentment, for example, might be theoretically irrational

given hard incompatibilism, but I maintain that all things considered they are suboptimal relative to alternative attitudes available to us. Second, the attitudes that we would want to retain either are not threatened by hard incompatibilism, since they do not have presuppositions that conflict with this view, or else have analogues that would not have false presuppositions. The attitudes and analogues that would survive do not amount to Strawson's objectivity of attitude, and are sufficient to sustain good interpersonal relationships.

Of all the attitudes associated with moral responsibility, moral resentment, that is, anger with an agent because of a wrong he has done to oneself, and indignation, anger with an agent because of a wrong he has done to a third party, seem most closely connected with it. It is telling that debates about moral responsibility typically focus not on how we should regard morally exemplary agents, but rather on how we should consider those that are morally offensive. The kinds of cases most often used to generate a strong conviction of moral responsibility involve especially malevolent wrong done to another, and the sense of moral responsibility evoked typically involves indignation. Perhaps, then, our attachment to moral responsibility derives partly from the role moral resentment and indignation have in our lives, and that hard incompatibilism is especially threatening because it challenges their rationality.

Let us combine indignation and moral resentment into a single notion, and call it *moral anger*. Moral anger is directed toward an agent who is represented as knowingly having done wrong, or as culpably negligent, and it presupposes that this agent, as a result, deserves in the basic sense to be a target of this sort of anger. Not all anger is moral anger. One type of non-moral anger is directed toward someone because his abilities in some respect are scant or in

some particular circumstance he performs poorly. We are sometimes angry with machines for malfunctioning. On occasion we are angry without any target, whereupon we might search one out. Still, most human anger is moral anger.

Moral anger forms an important part of human ethical life as it is ordinarily conceived. It motivates us to resist oppression, injustice, and abuse. But often expressions of moral anger have harmful effects. They often fail to contribute to the well-being of those to whom they are directed. Frequently expressions of moral anger are intended to cause physical or emotional pain. Partly as a result of these problems, moral anger often has a tendency to damage or destroy relationships. In extreme cases, it can provide motivation to take very harmful and even lethal action against another.

The sense that expressions of moral anger are damaging gives rise to a robust demand that they be morally justified. The demand to produce a moral justification for behavior that is harmful to others is always pressing, and expressions of moral anger are typically harmful to others. Moreover, this demand is made more acute by our attachment to moral anger; we often enjoy displaying it, and so we want these displays to be morally justified. We frequently justify expressions of moral anger by the claim that wrongdoers deserve in the basic sense to be their target. From the hard incompatibilist perspective, however, this claim is mistaken. Yet even if we sense that it is, we might still retain an interest in preserving the belief in moral responsibility to satisfy the demand to justify expressions of moral anger.

Accepting hard incompatibilism is all by itself unlikely to alter our psychology so that moral anger is no longer a problem for us. Nevertheless, much of such anger feeds on the

presupposition that its object deserves, in the basic sense, blame for a moral offense. Destructive anger in relationships is nourished by the belief that the other 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. Hard incompatibilism advocates retracting such beliefs because they are false, as a result of which the associated anger might be diminished, and its expressions reduced.

At the same time some types and certain degrees of moral anger are likely to be beyond our power to affect, and thus even supposing that the hard incompatibilist is committed to doing what is right and rational, she would still be unable to eradicate these attitudes. Shaun 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 hard incompatibilists we might expect that we cannot keep ourselves from some degree of narrow-profile, immediate moral anger. But in wide-profile cases, we might well have the ability to prevent, alter, or eliminate moral anger, and given a belief in hard incompatibilism, we might do so for the sake of morality and rationality. Modification of moral anger, aided by a hard incompatibilist conviction, might well be beneficial for personal and societal relationships.

But moral anger plays an important communicative role in relationships with others in both personal and societal relationships, and one might object that if one were to strive to modify or eliminate this attitude, such relationships might well be damaged. However, when one is wronged in a relationship there are other emotions typically present that are not

threatened by hard incompatibilism, whose expression can also communicate the relevant information. These emotions include feeling hurt or shocked about what the other has done, and moral sorrow, sadness and concern for the other. Parents might feel intensely sad, and not angry, that their son has driven his car while intoxicated and hurt a pedestrian, and they might be concerned that he will continue to behave in this way. Ordinary human experience indicates that communicating such sadness and concern can be an effective way to motivate avoidance of future bad behavior. Often communicating anger is at least not required to secure this effect, and frequently it is manifestly an inferior approach, given its deleterious consequences. Feigned moral sadness and sorrow are sometimes used to manipulate others, but what I have in mind are the genuine versions. These attitudes 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 anger is required, or optimal, for communication in interpersonal relationships.

Nichols argues that sadness together with moral resolve is an inadequate substitute for moral anger in personal and social relationships (Nichols 2007). His argument begins with the claim that moral anger can be shown, by way of empirical studies, to be beneficial to human beings in certain key respects. He then contends, also on the basis of empirical work, that sadness together with resolve 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 do not provide evidence that adult human beings, with education and determination, would not benefit overall from the substitutions in their personal and social relationships.

Nichols points out, correctly, that moral anger is effective in encouraging cooperation and signaling defection. In support, he cites a recent study in experimental behavioral economics:

> Ernst Fehr and colleagues have examined reactions to punishment in public goods games. A typical public goods game involves four subjects, playing anonymously on computers. Each subject is given an allotment of monetary units, and each is allowed to invest whatever portion he chooses into a common fund. For every 1 monetary unit an individual invests, 1.6 units go into the common fund, which is a net benefit for the group, but a net loss for the individual (since he only gets 40% of his investment back). Obviously it's optimal for the group if everyone invests in the common fund, but for each individual, it's selfishly better not to invest. Fehr & Gächter (2002) had subjects play a series of such games in which subjects were told (truly) that they would never interact with any player in more than one game. After each game, subjects were given an opportunity to pay to "punish" people in the group that they just played with; for each 1 monetary unit the punisher pays, 3 monetary units are deducted from the punishee's allotment. Remember that the subjects know that they will not play another game with any of these particular players, so punishing apparently has no future benefit for the subject. Nonetheless, punishment was common, and it was typically directed at

> defectors (i.e., individuals who contributed less than average) (Fehr & Gächter 2002, p. 137).
>
> In a striking extension of this work, Fehr & Fischbacher (2004b) explored whether external observers, "third-parties", would be willing to punish players. The third party observed two subjects in an economic game. Fehr & Fischbacher found that about half of the third-party participants paid to punish players who violated norms of cooperation. In this study, as well as Fehr & Gächter (2002), the motivation for punishment does not seem to be anything like explicit considerations about material gains—the punisher only loses money, and they have little reason to think the punishment will materially improve the situation of the other players.

Nichols then asks: “Why would subjects spend their money to punish even when it's obviously not in their material self-interest?”, and “anger” is the answer. He continues:

> Now, what are the consequences of this anger-driven punishment? As it happens, punishment dramatically affects behavior in these games. Fehr and colleagues have consistently found increased cooperation when punishment is an available option. Perhaps the most impressive illustration comes from experiments in which players first engage in several games in which punishment is not available. Fehr & Gächter (2000) conducted such an experiment in which subjects played for 10 rounds with no option of punishment; by the 10th round the level of contribution was quite low (below 20%). Then punishment was introduced as an option for the 11th round. Immediately, the

> contributions leapt to over 60% and within a few more rounds with punishment, the level of contributions was at 90%! Fehr & Gächter maintain that such punishment is driven by anger, and if they're right, then anger is a potent force for motivating cooperation (2002, 139).

Nichols draws our attention to three items concerning the relationship between punishment and cooperation in these studies. First, cooperation deteriorates in the absence of punishment. Many people start out contributing a significant amount of their endowment, but without punishment, this drops off dramatically. Second, the mere belief that punishment might be inflicted increases cooperation, and Fehr and Gächter argue that this is because people anticipate that if they defect, others will be motivated by anger to punish then. Third, punishment pushes cooperation close to the maximum. In summary, moral anger and its punitive expressions work to make cooperation highly likely by minimizing cheating, defection, and violations of reciprocity.

Let me begin with a clarification. My hard incompatibilist position is that punishment in the sense of death, suffering, and the degree of loss of liberty that imprisonment involves, inflicted on a person as a response to wrongdoing, should not be administered except insofar as it cannot be justified by analogy with quarantining carriers of dangerous diseases. Since property rights are not nearly as immune to being overridden by consequentialist considerations as the rights to life, liberty, and freedom from the infliction of significant pain, I see no objection to consequentialist justifications of monetary penalties for violation of traffic regulations and rules by which the economy functions (Pereboom 2001).

That said, I do not contest that co-operation can secured through fear of anger-motivated punishment, and that it can have this role in personal relationships, communities, and societies. However, I do think that it is not obvious that this is the optimal way of achieving cooperation (thanks to Mike Patterson for discussion of these issues). First, we prefer cooperation motivated by moral values, and not by the self-interested consideration of avoiding anger and punishment. Second, anger is a blunt instrument. It is very difficult to calibrate angry responses, such as anger-motivated punishment, optimally. An ethnic group is treated unjustly, angry responses result in punitive behavior, the punishment is far from fair, and anger motivates a new unfair regime of punishment. Familiarly, this pattern often occurs at the personal level. Third, anger-motivated punishment produces fear or even terror, and it is better, all else equal, that cooperation be motivated differently. There are further respects in which punishment is a sub-optimal method for securing cooperation, and examining the evolution of animal training is instructive on this issue (thanks to Sacha Sullivan for suggesting this parallel). Here is the testimony of one animal trainer:

> Most animal trainers have become champions of positive reinforcement and in doing so have evolved away from the use of force and aggression, the more traditional animal training tools that have been in use for thousands of years. Fortunately, negative reinforcement and punishment are slowly, often reluctantly, giving way to positive reinforcement in most animal training communities. The increased use of positive reinforcement has created opportunities that are beyond many people's imagination while at the same time creating better working environments for animals…. The

scientific community has demonstrated with hundreds of species from cockroaches to whales that the use of aversives, such as in negative reinforcement and punishment training strategies, produce certain detrimental side effects. These side effects include: aggression, escape/avoidance, generalized fear of the environment, and apathy or generalized reduction in behavior. Fighting the urge to use negative reinforcement and punishment is not always easy. Most people grow up in an environment where negative reinforcement and punishment were the tools that influenced their behavior. This cultural influence was, and still is, evident in a myriad of sources in everyone's lives. Parents, teachers, siblings, schoolmates, etc. all use a multitude of negative reinforcers (the threat of punishment) to force people to comply with wishes, rules and demands. They also punish people when they do not follow these rules or do not live up to certain expectations. Fear of punishment is a powerful motivator, but plagued with the detrimental side effects I mentioned above. The most effective motivator is, and has always been, positive reinforcement… The key is to understand that positive reinforcement is almost always the best training tool for the job, even when punishment or negative reinforcement may be easier and produce quicker results. Skilled animal trainers know that if they ever catch themselves stopping a behavior (punishment) they should immediately begin developing a plan for how to avoid using punishment in the future. Often this is accomplished by finding ways to use positive reinforcement to train a desirable behavior to replace the unwanted behavior…. [A]n animal trained with negative reinforcement or punishment will only perform at the

> level necessary to avoid the aversives, whereas an animal trained with positive reinforcement will look forward to the interactions with the trainer, will be more creative about how to earn the reinforcement and will experience far less stress than an animal trained with aversives. (Martin 2009)

One will find similar advice in virtually all contemporary animal training literature. A key point frequently emphasized is that expressions of anger and punitive responses very often have deleterious effects by comparison with non-punitive alternatives. Punishment causes fear, and those who are subject to threat of punishment will behave less creatively, for fear that if it behaves in an untested way it will be subjected to punishment. More generally, those who are subject to punishment perform at a lower level than those trained by alternative methods. Further, punishment and threat of punishment produces much more stress than positive training methods. In addition, the relationship with the trainer is better if punishment is avoided. Examining animal training is especially instructive, since often professional animal trainers are highly motivated to hinder anger or its expression if it will facilitate better performance. Also, the fact that it is possible for animal trainers to curtail anger and its expression counts as evidence we might also do so in our relations with humans.

My proposal is that moral sadness, sorrow, and concern, combined with resolve to effect salutary change is a superior way to secure exemplary behavior than is punishment motivated by moral anger. Nichols disagrees, for he does not believe that sadness, for example, is an adequate substitute for moral anger:

> Can sadness do the requisite work for moral anger, then? Well, to answer this we need to know more about sadness itself. The most important question concerns how sadness affects our behavior. For we know that moral anger produces behavior that discourages cheating, defecting, and mistreatment. What kind of behavior does sadness tend to produce? None, according to emotion theorists. Lazarus writes, "In sadness there seems to be no clear action tendency—except inaction, or withdrawal into oneself" (Lazarus 1991, 251). This is illustrated in infancy research. 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. 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). By contrast, infants who respond with anger are more likely to try to overcome the obstacle (Lewis & Ramsey 2005, 518). As a result, sadness seems too behaviorally weak to do the work of anger.

In response, the claim 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. Now 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. When a parent is too intensely sad to be angry about her son's drunk driving and hitting a pedestrian, experience indicates that in combination with resolve this attitude can motivate her take measures to cause him change his behavior. My claim is that for adults, sadness and sorrow, accompanied by a resolve to fairness and justice, or to improving one's personal relationships, will optimize personal and social relationships relative to punishment motivated by moral anger.

Nichols considers my proposal to supplement of sadness with resolve:

> Pereboom brings in a further element that might be thought to address the shortcomings of the analogue emotions: resolve. In addition to moral sadness, we might be committed to opposing wrongdoing, and this "would allow for a resolve to resist abuse, discrimination, and oppression" (Pereboom 2007, 124). In his discussion of guilt, he says something similar: "because you have a commitment to doing what is right, and to personal moral progress, you might resolve not to perform an immoral action of this kind again, and seek out therapeutic procedures to help treat one's character problems" (Pereboom 2001, 205). People no doubt differ in the strength of their resolve. But it strikes me as unlikely that resolve will provide sufficient motivation for the bulk of the population. After all, many teenagers think that they risk going to hell if they have sex, yet this often provides insufficient motivation for abstinence. Or consider the Marxist thought that working hard will generate benefits for the state which will in turn benefit everyone. This turns out to be motivationally feeble. Marxism seems incredibly rational, which is why it is so attractive to us intellectuals. But it turns

out to be naively optimistic about the plasticity of human motivation. I suspect the same is true of the revolutionary's hope for replacing problematic reactive attitudes. 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 is 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. At the societal level, a number of jurisdictions worldwide, such as Finland and the Netherlands, have altered their treatment of criminals so that it is much less punitive, and is instead more oriented toward public safety and rehabilitation, often with significant success. In raising children, more commonly than ever before, non-punitive methods have come to replace punishment, with good results. Again, modern animal training exhibits a similar evolution. One should also note that for many centuries now human beings have developed communities, within, for example, Buddhist and Christian traditions, in which training and teaching methods are employed to diminish moral anger, and to develop moral and religious excellence by other means. It's especially important that we examine such communities to see whether these alternative methods can be successful.

3. Blame.

The hard incompatibilist rejects the legitimacy of any blaming practice that presupposes that the agent being blamed is morally responsible in the basic-desert sense, or is an appropriate target of basic-desert entailing reactive attitudes. One might object that without the practice of blaming, moral improvement will be difficult to achieve. In response, there are notions of blame that will be acceptable to the hard incompatibilist (Pereboom 2008). In George Sher's analysis, blame is at its core a certain belief-desire pair; the belief that the agent has acted badly or is a bad person; and the desire that he not have performed his bad act or not have his current bad character (Sher 2006: 112). The hard incompatibilist can, without inconsistency, endorse these beliefs and desires about badness. It might be objected that if we gave up the belief that people are blameworthy, we could no longer legitimately judge any actions as good or bad. This isn't right. Even if we came to believe that a perpetrator of genocide was not morally responsible in the basic-desert sense because of some degenerative brain disease, we could still legitimately maintain that it was extremely bad that he acted as he did. In general, denying blameworthiness would not threaten judgments of moral badness, and, likewise, denying praiseworthiness would not undermine assessments goodness. So far, then, the hard incompatibilist can accept the legitimacy of blaming on Sher's analysis.

However, Sher also contends that blame involves a set of affective and behavioral dispositions, and at this point one might think his account to conflict with hard incompatibilism. But first, he does not regard any of these dispositions as essential to blame, but only connected to it in a looser sense. Given the looseness of this tie, the hard incompatibilist can endorse blaming in Sher's sense. She might not endorse all of the affective

and behavioral dispositions one might canvas – in particular, not those that presuppose or can only be justified in virtue of basic desert. Still, two important dispositions to which Sher draws our attention -- "to apologize for our own transgressions and vices and to reprimand others for theirs" (Sher 2006, 108), are fully compatible with a hard incompatibilist conviction.

In response to my position, Sher remarks:

> The deepest oddity about Pereboom's world, however, lies… in the fact that the only problems wrongdoing appears to present to its inhabitants are future oriented. That, at any rate, is the clear implication of the three responses to wrongdoing – admonish, ignore, walk away – that Pereboom is willing to countenance; for all three recommend themselves primarily as methods of preserving our future tranquility. This exclusively future-oriented stance toward wrongdoing, reminiscent of some of what Strawson says about the objective attitude, is bound to seem profoundly strange to anyone to whom the primary significance of wrongdoing lies not in what it augurs but simply what it is. (Sher 2006: 6)

But the hard incompatibilist does accommodate backward-looking attitudes toward wrongdoing that do not presuppose basic desert. These include sadness or sorrow about the wrongdoing of another, and, as we will see, regretting one's own wrongdoing. In addition, the essential elements of blame on Sher's account, which are backward-looking -- the belief that the agent has acted badly, or is a bad person, and the desire that he not have performed the bad act, or not have his current bad character – are also not undercut by any claim the hard incompatibilist makes.

In Thomas Scanlon's analysis, to blame an agent for an action is to judge that it shows 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 can be acceptable to the hard incompatibilist depends on how the appropriateness to which this characterization refers is construed. If this notion this is taken to introduce basic desert, then the result will be unacceptable. But there is an epistemic or evidential reading that accommodates hard incompatibilism. One of Scanlon's examples illustrates this idea. 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 you to take your relationship with him to reflect this diminished trust. Here the justification is at least partly, if not wholly, evidential. You believed Bill was trustworthy to a high degree, but you then acquired good evidence that he is not especially trustworthy, and thus a good reason to judge that an attitude of his is relationship-impairing. You now make this judgment, and take your relationship with him to be modified in way that it justifies as appropriate, that is, you take your relationship to be damaged because the bond of trust has been weakened. All of this is unobjectionable to the hard incompatibilist.

4. Guilt and repentance.

One might object that the self-directed attitudes of guilt and repentance are threatened by hard incompatibilism. There is much at stake here, the objector might claim, for

these attitudes are not only essential to good interpersonal relationships for agents prone to wrongdoing, but are also required for the moral improvement, development, and sense of integrity of an agent of this sort. Without the attitudes of guilt and repentance, such an agent would not only be incapable of reestablishing relationships damaged because he has done wrong, but he would also be barred from a restoration of his own moral integrity. For in the absence of the attitudes of guilt and repentance there are no human psychological mechanisms that can generate a restoration of this sort. Hard incompatibilism would appear to undermine guilt because this attitude essentially involves a sense that one is blameworthy, in the basic-desert sense, for what one has done. If an agent did not feel blameworthy for an offense, the objector continues, he would also not feel guilty for it. Moreover, because feeling guilty is undermined by hard incompatibilism, repentance is also no longer an option, since feeling guilty is required to motivate a repentant attitude.

However, suppose that you do wrong, but because you believe that hard incompatibilism is true, you reject the claim that you are blameworthy. Instead, you accept that you have done wrong, you feel deeply sad that you were the agent of wrongdoing, or as Waller advocates, you thoroughly regret what you have done:

> It is reasonable for one who denies moral responsibility to feel profound sorrow and regret for an act. If in a fit of anger I strike a friend, I shall be appalled at my behavior, and profoundly distressed that I have in me the capacity for such behavior. If the act occurs under minimum provocation, and with an opportunity for some brief reflection before the assault, then I shall be even more disturbed and disappointed by my

> behavior: I find in myself the capacity for a vicious and despicable act, and the act emerges more from my own character than from the immediate stimuli (thus it may be more likely to recur in many different settings), and my capacity to control such vicious behavior is demonstrably inadequate. Certainly, I shall have good reason to regret my character -- its capacity for vicious acts and its lack of capacity to control anger. (Waller 1990: 165-6)

Moroever, because you have a commitment to doing what is right, and to personal moral improvement, you would resolve not to perform an immoral action of this kind again, and perhaps seek out help to make this change. None of this is undercut by a hard incompatibilist conviction.

Nichols raises an objection to this proposal:

> What does the revolution have in store for guilt? Pereboom and Waller propose that regret can do substitute service for guilt. Pereboom writes, "suppose that you behave immorally, but because you endorse hard incompatibilism, you deny that you are blameworthy. Instead, you acknowledge that you have done wrong, you feel sad that you were the agent of wrongdoing, and you deeply regret what you have done" (Pereboom 2007, 120; see also Waller 1990, 165–7). Unfortunately, regret is not well-defined. Indeed, some theorists assimilate it to guilt (e.g. Storm & Storm 1987). Lazarus claims that the term is simply too ambiguous to be usefully compared to other emotions (1991, 244–5). Without a more detailed description of the emotion, it's hard to say whether it is likely to do the good services that guilt provides.

However, there is available a more detailed and precise characterization of regret, which we might alternatively designate as guilt without the presumption of basic desert. Hilary Bok eloquently describes this attitude:

> 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: 168-9)

Regret, or else guilt without basic desert, characterized in this way, is especially apt for motivating repentance, moral self-improvement, and taking steps to restore one's relationships. Blaming oneself in the basic-desert sense might well also achieve these results, but it is implausible that the attitude Bok describes would be less effective.

5. Forgiveness.

On one conception of forgiveness, this attitude presuppose that the person being forgiven deserves blame in the basic-desert sense, and if this is correct, forgiveness would indeed be undercut by hard incompatibilism. Dana Nelkin argues that forgiveness does not have this presumption, and I think she is right (Nelkin 2008). But even if it does, there are features of forgiveness that would not be jeopardized by the truth of hard incompatibilism, and they can adequately take the place this attitude usually has in relationships. Suppose a friend has wronged you in similar fashion a number of times, and you find yourself unhappy, angry, and resolved to loosen the ties of your relationship. Subsequently, however, he apologizes to you, which, consistent with hard incompatibilism, signifies his recognition of the wrongness of his behavior, his wish that he had not wronged you, and a genuine commitment to moral improvement. As a result, you change your mind and decide to continue the relationship. In this case, the feature of forgiveness that is consistent with hard incompatibilism is the willingness to cease to regard past wrongful behavior as a reason to weaken or dissolve one's relationship. Forgiveness of this sort can be viewed as retracting blame in the sense Scanlon characterizes it (Scanlon 2009). My forgiving someone who has wronged me would involve my having judged that what the other did showed something about his attitude toward me that impairs the relationship he can have with me, but due to his repentance, my no longer taking my relationship with him to be modified in a way that this judgment of impaired relations justifies as appropriate. The judgment of impaired relations is

withdrawn because I take the other to have given up the attitude toward me that impairs the relations he can have with me.

In another kind of case, I might, independently of the offender's repentance, simply choose to disregard the wrong as a reason to alter the character of our relationship. This attitude is also not called into question by hard incompatibilism. The sole aspect of forgiveness that is challenged by a hard incompatibilist conviction is the willingness to overlook blame or punishment deserved in the basic sense. If one has given up belief in such deserved blame and punishment, then the willingness to overlook them is no longer required for good interpersonal relationships.

6. Gratitude and love.

Gratitude might seem to presuppose that the agent to whom one is grateful is morally responsible in the basic-desert sense for a beneficial act, whereupon a hard incompatibilist conviction would undermine gratitude (cf. Honderich 1988: 518-9). But even if this is so, as in the case of forgiveness certain core aspects of this attitude would remain unaffected, and these aspects can provide what is required for good interpersonal relationships. Gratitude involves, first of all, being thankful toward someone who has acted beneficially. It is not implausible that being thankful toward someone usually involves the belief that she is praiseworthy for some action. Still, one can be thankful to a young child for some kindness without believing that she is morally responsible for it. This aspect of gratitude could be retained even without the presupposition of basic-desert praiseworthiness. Usually gratitude

also involves joy as a response what someone has done. But no feature of hard incompatibilism poses a threat to the legitimacy of being joyful and expressing joy when others are considerate or generous in one's behalf. Expressing joy can bring about the sense of harmony and goodwill often produced by a sense of gratitude unmodified by hard incompatibist belief.

Finally, one might contend that love between mature persons would be subverted if hard incompatibilism were true. Let us first ask whether loving another requires that she be free in the sense required for moral responsibility. One might note here that parents love their children rarely, if ever, because their children possess this sort of freedom, or because they freely (in this sense) choose the good, or because they deserve, in the basic sense, to be loved. Moreover, when adults love each other, it is also seldom, if at all, for these kinds of reasons. Explaining love is a complex enterprise. Besides moral character and action, factors such as one's relation to the other, her appearance, manner, intelligence, and her affinities with persons or events in one's history all might have a part. But suppose that moral character and action are of paramount importance in producing and maintaining love. Even if there is an important aspect of love that is essentially a deserved response to moral character and action, it is unlikely that one's love would be undermined if one were to believe that these moral qualities do not come about through free and responsible choice. For moral character and action are loveable whether or not they merit praise. Love of another involves, most fundamentally, wishing well for the other, taking on many of the aims and desires of the other as one's own, and a desire to be together with the other. Hard incompatibilism threatens none

of this.

One might contend, however, that gratitude and love that are themselves freely willed are genuinely valuable, and not worth nearly as much if they are not freely willed. Consider the following excerpt from John Milton's *Paradise Lost*, whose theme is a familiar topic of theological controversy:

> So will fall
> He and his faithless Progeny: whose fault?
> Whose but his own? ingrate, he had of me
> All he could have; I made him just and right,
> Sufficient to have stood, though free to fall. ..
> Not free, what proof could they have given sincere
> Of true allegiance, constant Faith or Love,
> Where only what they needs must do, appeared,
> Not what they would? what praise could they receive?
> What pleasure I from such obedience paid,
> When Will and Reason (Reason also is choice)
> Useless and vain, of freedom both despoiled,
> Made passive both, had served necessity,
> Not me. They therefore as to right belonged,
> So were created, nor can justly accuse
> Their maker, or their making, or their Fate;

As if Predestination over-ruled

Their will, disposed by absolute Decree

Or high foreknowledge; they themselves decreed

Their own revolt, not I: if I foreknew,

Foreknowledge had no influence on their fault,

Which had no less proved certain unforeknown.

So without least impulse or shadow of Fate,

Or aught by me immutably foreseen,

They trespass, Authors to themselves in all

Both what they judge and what they choose; for so

I formed them free, and free they must remain,

Till they enthrall themselves... (Milton 1667/2005, Book III, 95-125)

A key to Milton's vision of the meaning of the universe is that people have the opportunity to freely respond to God, and the freedom at issue is incompatible with causal determinism. Milton intimates that if divine grace were to causally determine responses such as gratitude and love, they would have little or no value; "Of true allegiance, constant Faith or Love/Where only what they needs must do, appeared/Not what they would? what praise could they receive?/ What pleasure I from such obedience paid/When Will and Reason (Reason also is choice)/Useless and vain, of freedom both despoiled/Made passive both, had served necessity/Not me.."

An idea suggested by Milton is that it is valuable to be loved by another as a result of

her free will, and that without free will having this role, love loses much of its value. However, against this, it is clear that parents' love for their children -- a paradigmatic sort of love -- is often produced independently of the parents' will. Robert Kane endorses this last claim, and a similar view about romantic love, but he nevertheless argues that a certain type of love we want would be endangered if we knew that there were factors beyond the lover's control that determined it:

> There is a *kind* of love we desire from others -- parents, children (when they are old enough), spouses, lovers and friends -- whose significance is diminished... by the thought that they are determined to love us entirely by instinct or circumstances beyond their control or not entirely up to them... To be loved by others in this desired sense requires that the ultimate source of others' love lies in their own wills. (Kane 1996: 88)

But setting aside *free* will for a moment, by contrast with voluntariness considered independently of freedom, in which types of case does the will intuitively play a role in generating love for another at all? When the intensity of an intimate relationship is waning, people sometimes make a decision to try to make it succeed, and to attempt to regain the type of relationship they once had. When one is housed in a dormitory with someone one didn't select, one might choose to make the relationship work. When one's marriage is arranged by parents, one may decide to do whatever one can to love one's spouse. In such situations we might desire that another person make a decision to love, but it is far from clear that we have reason to want the decision to be *freely* willed in the sense required for moral responsibility. A

decision to love on the part of another might greatly enhance one's personal life, but it is not obvious what value the decision's being free and thus praiseworthy would add. Moreover, while in circumstances of these kinds we might desire that someone else make a decision to love, we would typically prefer the situation in which the love was not mediated by a decision. This is true not only for romantic attachments, but also for friendships and for relationships between parents and children.

One might suggest that the will can have a key role in *maintaining* love over an extended period. Søren Kierkegaard suggests that a marital relationship ideally involves a commitment that is continuously renewed (Kierkegaard 1843/1971). Such a commitment involves a decision to devote oneself to another, and thus, in his view, a marital relationship ideally involves a continuously repeated decision. A relationship with this sort of voluntary aspect might in fact be highly desirable. Nevertheless, it is difficult to see what is to be added by these continuously repeated decisions being *freely* willed in the sense required for moral responsibility, as opposed to, say, expressing what the agent deeply cares about. Thus although one might at first have the intuition that freely-willed love is desirable, it is hard to see exactly where free will might have a desirable role in producing, maintaining, or enhancing love.

A worry might arise if the proposal to be evaluated is that the love causally determined by factors beyond one's control, for example, as in Milton's imagined scenario, by God. For a striking case, one Milton has in mind, would love for God be valuable to him if he causally determined us to love him? Still, perhaps even then only the specific character of the causal

determination might be objectionable. Suppose Ann causally determines you to love her by manipulating your brain so that you are oblivious to her flaws of character, and by slipping *Love Potion Number 9* into your morning coffee. That would be objectionable. But imagine instead that you have a self-destructive proclivity to love people who are harmful to you, and not to love those who would benefit you, partly because you have a tendency overlook people's valuable characteristics, such as kindness and concern for the well-being of others. Suppose Ann slips a drug into your coffee that eliminates this tendency, as a result of which you are able to fully appreciate her valuable characteristics, and as a result you are causally determined to love her. How bad would that be? It would seem that what is unacceptable is not being causally determined to love by the other party *per se*, but rather how one is causally determined, and that there are varieties of determination by the other party that are not objectionable.

7. Final words.

I conclude that we lack sufficient reason to believe that thinking and acting in harmony with hard incompatibilist convictions would on balance hinder personal and societal relationships. The hard incompatibilist would resist moral anger and punishment motivated by this attitude, but she would not be exempt from sorrow or sadness upon being wronged, and with moral resolve she would be motivated to take the positive measures required to restore and improve her relationships. When hurt by another, she might blame in the senses that Sher and Scanlon specify, since they do not presuppose basic desert, and upon acknowledgment of

wrongdoing the other's part, cease to regard it as an obstacle to her relationship. She would be thankful and express joy toward others for the good things they provide for her. Her beliefs pose no obstacle to love. If she had a independent tendency to control or distance herself from another she might come to see him "as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided..." (Strawson 1962). Her hard incompatibilist conviction, however, would not motivate or justify assuming this objectivity of attitude.

References

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

Fehr, E., Fischbacher, U. 2004a. "Social Norms and Human Cooperation," *Trends in Cognitive Sciences* 8: 187–90.

Fehr, E., Fischbacher, U. 2004b. "Third Party Punishment and Social Norms," E*volution and Human Behavior* 25: 63–87.

Fehr, E., Gächter, S. 2000. "Cooperation and Punishment in Public Goods Experiments," *American Economic Review* 90: 980–94.

Fehr, E., Gächter, S. 2002. Altruistic Punishment in Humans," *Nature* 415: 137–40.

Honderich, T. 1988. *A Theory of Determinism*, Oxford: Oxford University Press.

Hume, D. 1739/1978. *A Treatise of Human Nature*, Oxford: Oxford University Press.

Kane, R. 1996. *The Significance of Free Will*, New York: Oxford University Press.

Kierkegaard, S. 1843/1971. *Either/Or*, v. 2, tr. Walter Lowrie, Princeton: Princeton University Press.

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

Martin, S. 2009. "The Art of Training," *Natural Encounters*, http://www.naturalencounters.com/images/Publications&Presentations/The_Art_Of_Training-Steve_Martin.pdf)

Milton, J. Paradise Lost. 1667/2005. W. W. Norton and Company.

Nelkin, D. 2008. "Responsibility and Rational Abilities: Defending an Asymmetrical View," *Pacific Philosophical Quarterly* 89, pp. 497-515.

Nichols, S. 2007. "After Compatibilism: a Naturalistic Defense of the Reactive Attitides," *Philosophical Perspectives* 21, pp. 405-28.

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. 2008. "Defending Hard Incompatibilism Again," *Essays on Free Will and Moral Responsibility*, Nick Trakakis and Daniel Cohen, eds., Newcastle: Cambridge Scholars Press: 1-33.

Russell, P. 2000. "Compatibilist-Fatalism", in Ton van den Beld, editor, *Moral Responsibility and Ontology,* Dordrecht: Kluwer, 2000), pp. 199-218.

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

Sher, G. 2006. *In Praise of Blame*, Oxford: Oxford University Press.

Sommers, T. 2007. "The Objective Attitude," *The Philosophical Quarterly* 57: 321-41.

Sommers, T. 2009. "More Work for Hard Incompatibilism," *Philosophy and Phenomenological Research* 79 (2009), pp. 511-21.

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

Spinoza, B. 1677/1985. *Ethics*, in *The Collected Works of Spinoza*, ed. and tr. Edwin Curley, Volume 1, (Princeton: Princeton University Press).

Storm, C., Storm, T. 1987. “A Taxonomic Study of the Vocabulary of Emotions,” *Journal of Personality and Social Psychology* 53: 805–816

Strawson, P. F. 1962. “Freedom and resentment,” *Proceedings of the British Academy* 48: 1–25.

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