*Why We Are Responsible for Our Emotions**

EUGENE SCHLOSSBERGER

It is often said that one cannot be held responsible for something one cannot help. Indeed, Ted Honderich, Paul Edwards, and C. A. Campbell have suggested that it is obtuse, barbaric, or a solecism to think otherwise.[1] Thus, if (*contra* Sartre and others) one cannot help feeling one's emotions, one is not responsible for one's emotions.

In this paper I will argue otherwise; one is responsible for one's emotions, even if one cannot help feeling them.[2] In particular, I will define a rather special sense of the word 'responsible', one that is closely tied to our ordinary notion of moral worth. We are, in that special sense of the word, 'responsible' for our emotions. I will then argue that my limited sense of 'responsibility' is sufficient to warrant moral evaluation of individuals, and can be used as the basis for a theory of punishment.

1. *The strict sense of 'responsibility'*

There are many ways in which the word 'responsibility' is used in English. I wish to define a rather special sense of the word. When I say that A is morally responsible for x, I will mean simply that x is the sort of thing that moral theories evaluate, deem good, bad, or indifferent, and that any moral taint or lustre possessed by x applies as well to A. Thus there are two conditions for moral responsibility.

Firstly, it must be appropriate to ask whether x is morally offensive or desirable; x must be proper grist for the moral mill. It does not follow that x must be either good or bad. One would expect, for example, that a good moral theory would hold that choosing skiing over reading detective stories as a form of recreation is morally neutral (that is, it is neither desirable nor undesirable). Yet such a choice does come under the purview of moral scrutiny. A utilitarian, for example, might find that in opting for the slopes

* An earlier version of this paper was presented to the Western Division of the American Philosophical Association in April 1984. I would like to thank Myles Brand, Joel Feinberg, Peter van Inwagen, the editor of *Mind*, and others too numerous to mention for their helpful comments, and the NEH for its kind support.

[1] See, respectively, 'One Determinism', in *Essays on Freedom of Action*, Routledge and Kegan Paul, 1973, 'Hard and Soft Determinism', in Sidney Hook, ed., *Determinism and Freedom*, New York University Press, 1958 and 'Is Free Will a Pseudo-Problem?', *Mind*, 1951, pp. 441–65.

[2] P. S. Greenspan (in 'Behavior Control and Freedom of Action', *Philosophical Review*, 1978, pp. 225–40) distinguishes between what is within an agent's control, what is explicable in terms of rational decisions, and the ability to have done otherwise. My arguments concerning control apply, *mutatis mutandis*, to explication in terms of rational decisions and the ability to do otherwise.

I have done neither badly nor well, since my skiing and my reading Agatha Christie would produce, overall, equal amounts of happiness. The important point is that my decision to go skiing is found to be morally neutral as a result of a *moral* calculation. Similarly, a Kantian might seek an appropriate maxim, and ask whether the maxim expressed by my decision is coherently universalizable. In both cases there are criteria for moral evaluation that apply to my choice.

Moral theories, in other words, establish moral parameters; they indicate which aspects of the human condition have moral significance. A strict utilitarian, for example, holds that the only moral parameter is happiness. He would regard all and only those things that lead to happiness or unhappiness as subject to moral scrutiny. Alternatively, certain forms of Christianity hold that the only moral value is obedience to God. For such a view, x is appropriate for moral scrutiny if x can be meaningfully evaluated in terms of conformity to God's will. In general, a fully developed moral theory defines a set of traits, acts, events, intentions, etc., that are appropriate objects of moral evaluation.

It follows that appropriateness for moral evaluation depends upon the moral theory used to make the evaluation. Since we are primarily concerned with praising and blaming, however, the matter is somewhat simplified. In praising or blaming someone, one generally draws upon (explicitly or implicitly) a moral theory. So when I say 'x is appropriate for moral scrutiny', I will mean that the moral theory employed in evaluating A has criteria for evaluation under which x falls, that x belongs to the set of things that are appropriate objects of moral evaluation, according to the moral theory used to evaluate A. Of course, if one wants one's evaluation of A to be 'correct', one must use the 'correct' moral theory. If there is an objectively correct moral theory, that is the theory to be used in determining whether or not x is appropriate for moral scrutiny. If there is no correct moral theory, then the theory that determines whether x is a proper object for moral scrutiny is the theory used to evaluate A.

It follows that in order to make substantial claims about responsibility, I must assume some moral theory (or at least range of moral theories). For the sake of simplicity I will assume that the relevant moral theory is not overly bizarre (non-standard). In addition, I will rely upon a few widely though not universally shared moral assumptions.[3]

The second condition for moral responsibility (in my restricted sense) is that whatever moral taint or lustre x possesses attaches to A. To the extent that x is morally offensive or illustrious, A is morally offensive or illustrious. For example, although Cain's killing of Abel is morally pernicious, *my* moral worth is not affected by the deed. So I am not morally responsible for Cain's

[3] For example, as Myles Brand has pointed out to me, my assumption that enjoying the suffering of animals is the sort of thing that bears moral evaluation is not warranted by certain moral theories, e.g. pure hedonistic egoism.

killing of Abel. On the other hand, I am responsible for having chosen to ski rather than read detective stories. After all, were choosing to ski morally offensive, *my* choosing to ski would render *me* (to some extent) morally odious. In short, I am morally accountable for my choice, even though, as a matter of fact, I incur no moral credits or debits.

Now this account of responsibility clearly differs from the way some philosophers have used the word 'responsibility'. And perhaps Ted Honderich is right that *ordinarily* when we say that someone is responsible for something, we presuppose that he could have done otherwise.[4] But I am not trying to give an account of our ordinary notion of responsibility. Indeed, I would suggest that the more restricted notion should, at least in a number of important contexts, *replace* the ordinary notion.[5]

However, it should be said that the limited sense of 'responsibility', i.e. moral evaluability, is not so far removed from the 'ordinary' sense as is sometimes thought. Richard Brandt points out that

> When philosophers say that human beings are 'morally responsible' for their actions, what they apparently mean . . . is that it is right and proper, sometimes, to engage in blaming and praising . . . People are sometimes fittingly, deservedly, praised and blamed.[6]

Now according to the *Oxford English Dictionary*, to 'blame' is 'to find fault with, to censure'.[7] Thus to find moral fault with A is to (morally) blame him, and hence to find him 'morally responsible' in Brandt's sense. Thus if A has a quality that renders greater or less his moral value, he is 'morally responsible' for that quality in at least one 'ordinary' use of the term. And it is undeniable, I think, that we often make use of the notion of a trait or deed's affecting an agent's moral worth, whether or not we are tempted to use the word 'responsibility' to express that relationship. It is in this sense of the term that I claim that persons are 'responsible' for their emotions.[8]

[4] Ted Honderich, 'One Determinism', in *Essays on Freedom of Action*, Routledge and Kegan Paul, 1973.

[5] In particular, I would argue that the limited sense of 'responsible' is all that is presupposed by punishment, contract law, moral assessment of the worth of individuals, and is enough to support Strawson's 'essentially personal' human interactions.

[6] Richard Brandt, 'Determinism and the Justifiability of Moral Blame', in Sidney Hook, ed., *Determinism and Freedom*, New York University Press, 1958, p. 149.

[7] *Oxford English Dictionary*, Oxford, 1933.

[8] My remarks in this section are not intended to provide an *account* of moral responsibility, but merely to explain what I mean by the term. A variety of apparently odd consequences of my explanation disappear when the full account is introduced. Put briefly, I hold that one is responsible for evidencing sufficiently specific traits (ascribable features) that express one's worldview (that is, the set of beliefs, attitudes, dispositions, etc., both specific and general, transitory and enduring, that together constitute one's way of understanding and responding to the world about one). More precisely, A is responsible for a trait x iff:

(1) x includes all information relevant to making a moral evaluation (according to the moral theory being employed),

(2) x is partially constitutive of personhood (it expresses A's worldview), and

2. *Emotions*

Consider the case of Charlie. Charlie derives great pleasure from the sight of animals in pain. Mind you, he never *causes* animals to suffer. But when he chances upon a wounded animal, the sight affords him much delight.

Now recall that all it means to say that Charlie is responsible for relishing the sight of animals in pain is that (1) enjoying the sight of animals in pain is the sort of thing that the relevant moral theory evaluates (looks upon with moral favour, disfavour or neutrality), and (2) that to the extent that such enjoyment is a good or bad thing, Charlie is morally illustrious or flawed. Both conditions, I will argue, are met.

It seems clear that the first condition is met. It may be that enjoying the pain of animals is morally neutral. But it is hardly inappropriate to inquire about the moral desirability or undesirability of such feelings. After all, if concern for others is morally valuable, then feelings that reflect indifference to the sufferings of others are morally undesirable. Surely it would be preferable, from a moral standpoint, were such enjoyment not to occur. (Even a utilitarian, who values happiness, might hold that valuing happiness in others is preferable, morally speaking, to not valuing happiness in others.) In any case, it seems clear that delight in the suffering of others is the sort of thing to which a moral calculus can be applied; one's moral theory ought to take a stance on Charlie's feeling. And that is all that is required for condition (1) to be met.

It seems equally odd to deny that the second condition is met. To the extent that taking pleasure in the pain of animals is morally unsavoury, Charlie is morally unsavoury. After all, if concern for the well-being of others is morally valuable, it seems impossible to deny that it would be a moral improvement were Charlie to stop enjoying the sight of animals in pain. Again, surely someone who is otherwise just like Charlie but does not enjoy the torments of animals is morally preferable to Charlie. So Charlie would be a morally better person were he to stop taking such sadistic pleasure. And hence it seems absurd to claim that although Charlie's feeling is morally odious, and that although Charlie would be a better person were he to stop having that feeling, Charlie himself is in no way morally flawed.

So Charlie is responsible for enjoying the sufferings of beasts even though his conduct is faultless, and even if he cannot help feeling the way he does. For nothing in the above argument requires that Charlie be able to control his emotion. So even if he can, in fact, control his emotion, his responsibility for that emotion does not rest upon his ability to control it.

(3) A has the status of a full moral person (he or she has a sufficiently developed worldview that is reflected in his or her conduct).

Thus if Joshua's son has committed a murder, Joshua may be responsible for instantiating the trait of feeling paternal love for a murderer, but not for instantiating the trait of having committed a murder (for he did not instantiate that trait). Time does not permit a full treatment of this account of responsibility and its application to problematic cases.

At this point the reader may wish to voice any of several objections. Firstly, he might say, this result is counterintuitive; it seems wrong to hold someone responsible for his emotions. How can we blame Charlie, particularly if he never acts badly, never tries to cause animals to suffer? Secondly, he might object that my claim violates the dictum that ought implies can. Thirdly, he might insist that we can control our emotions; he may say, with Sartre, that we can choose our emotions, or he might say that we can do various things to bring about a change in our emotional responses. Finally, he might agree that we are responsible for our emotions in my restricted sense of 'responsibility'. But it is not that restricted sense with which we ought to be concerned. I will answer each of these in turn.

It goes without saying that these do not exhaust the possible sources of reluctance to accept my conclusion. Indeed, some reasons for thinking that control is a prerequisite for blame, such as the argument from moral luck, can be properly addressed only in the light of a full account of moral responsibility such as the one sketched briefly in n. 8. It is enough, for the present, to establish a prima-facie case. The burden of proof now rests upon anyone who would deny my conclusion.

The principle of alternate possibilities. The charge of counterintuitiveness rests upon two foundations. Firstly, my view denies a widely shared pre-philosophical conviction that we cannot be held responsible for emotions we cannot help feeling. Secondly, my view seems to have several counter-intuitive consequences. Neither charge, I will argue, provides a good reason for rejecting my view.

No doubt it does seem counterintuitive to say that Charlie is responsible for his emotions even if he cannot help having them. Many subscribe to the 'Principle of Alternate Possibilities' (hereafter abbreviated as PAP), i.e. the view that an agent cannot be responsible for x unless he or she could have done something other than x. But a bit of reflection may help to soften our prejudice.

The first point to remember is that there are several legitimate uses of the term 'responsible'; I am only claiming that Charlie is responsible for his emotions in one of these senses. There may be other senses in which he is not responsible. My claim is just that Charlie may be judged adversely, deemed morally defective, because of his emotion.

And it is not so odd to say that Charlie's emotion makes him a morally worse person. For the fact that Charlie enjoys the sight of an afflicted animal does, after all, reveal something about him as a moral being. Whether or not they are chosen, one's emotions express and partially embody one's values, one's moral outlook upon the world. If, for example, one feels anger at the deeds of terrorists, one does so (generally) because one has values that are violated by those acts, and anger seems the appropriate moral response. (Indeed, one might hold that someone who does not feel anger at such

injustice is morally flawed.) Similarly, if Harold feels pleasure when an attractive stranger smiles at him while walking down the street, it is not because Harold is fond of lip movements. There is nothing in the lip movement, *per se*, to make Harold feel happy. Rather, Harold feels pleasure because, for example, the smile reinforces Harold's belief that he, Harold, is a handsome fellow, and because Harold has certain attitudes about the value of physical attractiveness, or about the value of things one can expect to obtain because one is attractive. Thus Harold's emotion, like Charlie's, draws upon, expresses, is shaped by, and embodies a complex network of attitudes, values, aspirations, expectations, etc. What Charlie's emotion reveals is that there is something morally offensive about the values and attitudes Charlie holds. And surely someone who has morally invidious values is a morally flawed person. Since salubrious values are morally better than invidious values, I cannot see how one can deny that Charlie would be morally better were he to exchange his invidious values for salubrious ones. Thus to the extent that Charlie's emotions are morally invidious, Charlie is morally invidious. And that is all I mean by saying that he is 'responsible' for his emotions.

Indeed, it seems bizarre to say that although Charlie's feelings are morally unsavoury, Charlie is in no way morally unsavoury. To say this is, it would seem, to deny that Charlie's emotions are part of what he is, to make Charlie's emotions external to him; just as an ugly hat can be worn by a beautiful person, the traditional view seems to be saying, so a good person can have evil emotions. But this view of Charlie is surely mistaken. His feelings, after all, are not external appendages with which he finds himself burdened; they are not, as it were, lint clinging to his moral agency. Charlie is not some mythical 'pure chooser' devoid of emotions who is reduced to choosing between the buttons on some 'inner console', a moral ghost whose choices are limited by the emotions, aptitudes, and attitudes his console includes. Charlie is not a homunculus inhabiting the nether regions of Charlie's skull who cannot be faulted because the 'inner console' at his disposal does not contain the button 'do not feel this emotion'. Rather, Charlie's emotions are part of the person he is, part of his moral being. And if Charlie's emotions are invidious, then to that extent Charlie is invidious.

Moreover, a closer look at PAP reveals some difficulties worth pointing out. The arguments that follow do not *refute* PAP. But they do raise enough of a doubt to suggest that compatibilists and soft determinists, in accepting PAP, have granted too much to their adversaries. Rather than argue, as most compatibilists have done, that PAP is compatible with determinism, they ought to deny PAP.

Firstly, since the main argument for PAP is its purportedly universal acceptance, it is worth pointing out that PAP has come under attack in recent years. Daniel Dennett points out that we never inquire whether a person could have done otherwise before praising him or her. Moreover,

if 'responsibility really did hinge' on PAP, argues Dennett, 'it would be unlikely in the extreme . . . that anyone would ever know whether anyone has ever been responsible.' For 'it is possible for all we know that our decisions and actions truly are the magnified, macroscopic effects of quantum-level indeterminacies occurring in our brains.'[9]

Even more telling are the 'Frankfurt' counterexamples, the first of which appeared in Harry Frankfurt's 'Alternate Possibilities and Moral Responsibility'.[10] A particularly powerful variation runs as follows.

Suppose that Harry intends to slap his landlord at 12 p.m. He begins to do so. At that moment, unbeknownst to Harry, William presses a button on his machine that first blocks Harry's neural impulses and then swings Harry's arm, resulting in a resounding slap. Harry, who mistakenly believes that the slap was caused by neural impulses generated by his brain, is quite satisfied. Now it is clear that Harry could not have done otherwise than slap his landlord. In fact, the cause of the slap was external, thus violating Aristotle's criterion that the cause be internal to the agent, as well as the causal interpretation of Hume's criterion that the action 'proceed' from 'some cause in the character and disposition' of the agent. Harry did not cause the slap; William did. But however much we wish to blame William, do we really want to say that Harry is not responsible? Must Harry's landlord, if he is to be reasonable, absolve Harry of guilt? That would be a case of extraordinary moral luck. For it seems evident beyond dispute that Harry is no less culpable and no less morally answerable for the slap than he would have been had William not pressed the button. Perhaps the landlord could not recover damages in a civil suit. But Harry's moral slate is just as dirty.

Perhaps Harry is responsible only for attempting to slap his landlord. Usually, however, we use the language of attempt to indicate that the intended result did not occur. And in Harry's case the intended arm movement did occur. Nothing of moral significance distinguishes Harry's case from ordinary slappings. If Harry is responsible only for an attempt, then the rest of us are responsible only for attempts to act.

Harry's case demonstrates two points. Firstly, it is at least plausible to hold Harry responsible. Even those who would absolve Harry must recognize the existence of contrary intuitions. Hence PAP cannot be considered unproblematically true; PAP, in other words, must be argued for, not merely invoked as a truism. Secondly, one must be circumspect in describing what an agent is responsible *for*. I will return to this point shortly.

Similar counterexamples can be given to variants of PAP advanced since Frankfurt's paper. Consider, for example, Peter van Inwagen's 'prevention principles.'[11] Van Inwagen distinguishes between two principles. PPP1 asserts that 'a person is morally responsible for a certain event-particular

[9] 'I Could Not Have Done Otherwise—So What?' *Journal of Philosophy*, 1984, p. 557.
[10] *Journal of Philosophy*, 1969, pp. 829–39.
[11] Peter van Inwagen, *An Essay on Free Will*, Oxford University Press, 1983.

only if he could have prevented it' (p. 167). In applying this principle, van Inwagen uses a modification of Davidson's criterion for event-particular identity; x and y are the same event-particulars if and only if x and y have the same causes. PPP2 maintains that 'A person is morally responsible for a certain state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining' (p. 171). Both principles, according to van Inwagen, are immune to Frankfurt-style counterexamples.

Now PPP1 is indeed immune to Frankfurt-style counterexamples, but only because, given van Inwagen's criterion for being the same act, I would not be responsible for having done any act x given *any* significant counterfactual variations. For example, even if determinism is false, I would not be responsible for having done x had I acted from a different motive. For then the causes of my action would have been different, and so the action I would have performed would not have been x, however similar to x it might have been.

What this discussion shows, I suggest, is that, strictly speaking, we are responsible not for acts as such but for the properties we instantiate by performing those acts; the proper objects of judgments of responsibility are (sufficiently specific) traits or properties, not actions as such.[12]

Suppose, for example, that Mickey tells a lie under threat of immediate execution should he fail to do so. Mickey is responsible for instantiating the property of *having chosen to lie rather than be executed.* In addition, Mickey instantiated various other properties, such as *having spoken with a firm and steady tone in the face of danger*, for which he bears responsibility, although not blame. By contrast, *having chosen to lie* is not a sufficiently contextualized property, for it leaves out information required by the moral theory used to evaluate Mickey.

This approach has obvious pragmatic advantages. Since a single action involves numerous traits, it is possible to give a more flexible and precise account of an agent's moral standing by evaluating each of his or her traits than it is by making a judgment about a single action. For example, suppose that Abner believes that all non-Christians will suffer eternal torment unless killed with a silver knife, and so, after his best attempts at suasion have failed, kills Abgorrah with an argent blade. Abner may be blamed for instantiating the property of having violated someone's right to life, praised for instantiating the property of having acted from concern for another's well-being, blamed for his fanaticism and praised for his commitment to conviction, and so forth. One might hold him responsible for acting without sufficient epistemological warrant, but not for having held a false belief. The analysis can be as complex and multifarious as one's moral theory permits. By contrast, the traditional view permits only a single question: 'was Abner responsible for killing Abgorrah?'

[12] Van Inwagen himself, it should be said, prefers PPP2 to PPP1, since he thinks we are responsible not for event-particulars but for (generic) states of affairs.

Moreover, taking traits or properties instantiated to be the bearers of responsibility permits a more satisfactory treatment of omissions such as *not having given to charity for twenty years.* For while actions take place in a given time and a given place, and are tied to particular bodily movements, traits may be instantiated even when no particular deeds are being performed. (One may instantiate the trait of being strong even while supine with fatigue.)

Still, the claim that we are responsible not for actions as such but for properties will strike some as counterintuitive. I do not think it is. After all, no major moral theory evaluates actions directly; they all evaluate actions via specific properties that are of moral interest. A finger movement is not, *qua finger movement*, of moral significance. What is significant about that finger movement is that in making it one instantiates some morally significant properties (for example, being a murderer). But if the basis of moral evaluation is not my actions as such but the properties I instantiate, it seems natural, intuitively correct, to hold me accountable for instantiating those properties, and not for my actions as such. If what is of moral significance about my finger movement on the trigger is the malicious disregard for human life it evidences, then what I am morally accountable for is that malicious disregard, not the finger movement by which I evidence it. Thus our habit of speaking of responsibility for actions is *merely* a habit, and does not suit well our deeper conceptions of moral evaluation.

If I am right that we are responsible for properties instantiated rather than acts as such, it follows that PPP1 is true for the trivial reason that we are *never* responsible for event-particulars as such. Hence PPP1, though true, does not establish the necessity of alternative possibilities. What van Inwagen must argue for is not PPP1, but rather the claim that we are not responsible for instantiating a property P unless we could have failed to instantiate P. And this principle *is* susceptible to counterexamples. A brief example will suffice. Suppose that William decides to give me a particular set A of beliefs, attitudes, dispositions, etc. He performs neurosurgery, altering my brain so that I have set A. Now by chance I happened to have set A before William began his procedure. Thus my brain after the procedure is no different than it was before William began. As a result, my subsequent actions are no different than they would have been without the procedure. I am just as susceptible to rational persuasion as I was before, etc. Does William's procedure make any difference to my moral status? Am I henceforth absolved of all moral responsibility?

PPP2 raises similar problems. Van Inwagen imagines that A witnesses a brutal assault occurring outside her window, but decides not to call the police. Unbeknownst to her, her telephone is out of order. Van Inwagen claims that A is responsible, perhaps, for *deciding* not to call the police, but not for failing to call. And she is certainly, avers van Inwagen, not responsible for the fact that the police were not called.

Van Inwagen's claim is not as unproblematic as he seems to think it is. Suppose, for example, that a law, known to A, requires citizens witnessing an assault to call the police. Suppose as well that the *mens rea* requirement for this crime of non-reportage is that it be committed intentionally. Are we to say that A is not guilty, either legally or morally, of violating the law of non-reportage, that she is morally or legally guilty only of *attempted* non-reportage?

The drawbacks of holding individuals responsible for actions now become evident. If instead we hold individuals responsible for (sufficiently specific) properties they instantiate, such difficulties do not arise. A is morally responsible for the disregard for the law evidenced by her non-reportage. For whatever the *causes* of her failure to call the police, her inaction did *evidence* disregard for the law.

But now the inability to have done otherwise does not seem to obviate A's responsibility, except in so far as inability counts as *evidence* that A did not fail to take her legal duties seriously.

These cases are not mere technical difficulties for PAP and its variants. What they show is that we are not really concerned, when making moral evaluations, with the *causes* of an agent's actions. We are concerned rather with the moral character of the agent. As a rule, when a person has no alternative to doing x, his or her doing x reveals little about his or her moral character. Thus PAP has a certain surface appeal. But the appeal is illusory. For although we often seem to appeal to inability as an excuse, the inability is not itself the grounds for exculpation, but merely an indication that some other exculpatory conditions obtain. For example, suppose that I say 'don't blame me for wrecking your car; I couldn't help it. The brakes failed.' We recognize this as a valid excuse. But its validity rests not so much upon the fact that I could not have done otherwise as it does upon the fact that no attitude, belief, value, or morally significant disposition of mine is revealed by or reflected in the damage done to your car.

In sum, one cannot deny responsibility for one's emotions merely by invoking PAP. For there are good reasons to think that PAP is incorrect. Even if the above considerations do not *prove* the falsity of PAP, enough has been said to shift the burden of proof upon the holder of PAP.

Resisting emotions. None the less, one might charge, my view has some counterintuitive consequences. Consider the case of Robert, who not only enjoys the sight of animals in pain, but also tortures them in order to see them suffer. Surely Robert is not in the same moral position as Charlie. Yet both equally enjoy the sight of animals in pain. Indeed, one might even think Charlie praiseworthy, not blameworthy, for resisting the temptation to torture animals. Charlie cannot help feeling what he does (let us suppose), but he can control what he does about it. And it is that, the objection runs, for which he should be held responsible.

My response is that, according to my view, Charlie is indeed better than Robert. For, given my account of responsibility (as supplemented by n. 8), Charlie is responsible for instantiating two traits, the trait of having felt enjoyment in the suffering of animals and the trait of refraining from torturing animals. The former is blameworthy and the latter praiseworthy. Indeed, it may be that the praise due him for the latter trait outweighs the blame due him for the former. It may even be the case, depending upon one's moral theory, that Charlie is more praiseworthy than Ellie, who does not torture animals and never felt any pleasure in their suffering. He is certainly less blameworthy than is Robert, who instantiates the particularly evil trait of having tortured animals. Thus my view can account for the difference between Robert and Charlie; it is even consistent with (though it does not entail) the claim that Charlie is a better person than he would be if he never felt the urge to torture animals. (This is the case if one's moral theory values overcoming temptation more than never feeling temptation.) So this objection does not tell against my view.

Ought and can. Some readers will, no doubt, be tempted to object that my claim runs afoul of the dictum that 'ought' implies 'can'; it is, they would hold, illicit, senseless, or improper to say that A ought to do x if A cannot do x. Thus A cannot be held responsible for not doing x if he cannot do x.

This objection, I would hold, conflates exhortation with evaluation. It is indeed pointless to exhort someone to do what he cannot do, senseless to direct someone to perform an impossible deed. Since the word 'ought' is often used to direct behaviour, there is indeed a sense in which 'ought' implies 'can'. But it does not follow that one cannot be evaluated on the basis of things one cannot help. For to insist that moral goals and values are exhausted by what one can sensibly be exhorted to do or be is to beg the question. Even if 'A can sensibly be exhorted to do x' entails 'A can do x', it does not follow that 'A would be a better person were he x' entails 'A can be x'. (Indeed, to conflate duty with moral value is to deny meaning to the concept of supererogatory acts.)

An analogy might prove helpful. A physically underdeveloped person, let us suppose, can never be a good football player. It is bootless to exhort him to play well; he cannot. It does not follow, however, that it is improper to evaluate his play. He engages in playing football, and is a bad football player, whether or not he can help being so. Similarly, although it may be pointless to exhort one not to feel a given emotion, one who feels a morally offensive emotion can be deemed morally offensive.

Controlling emotions. Some philosophers have argued that we are responsible for our emotions because they are within our control. Thus, they might argue, the fact that Charlie is a worse person because of his unsavoury emotion does not show that we are responsible for things beyond our control.

This objection misconstrues the nature of my argument. For my argument in this paper does not purport to show that we are responsible for things that are, in fact, beyond our control (although I believe we are). Rather, it is meant to show that responsibility does not depend upon control. That is, I am not arguing that

(1) We are responsible for our emotions
(2) We cannot control our emotions
(3) Therefore we are responsible for things beyond our control.

Rather, my argument is

(1) We are responsible for our emotions *even if* we cannot control them
(2) Therefore controlling our emotions is not a precondition for being responsible for them
(3) Therefore control is not a precondition for responsibility.

That is, my argument that we are responsible for our emotions does not depend in any way upon our being able to control our emotions. Hence control is not, *per se*, a requirement for moral responsibility. So my claim here is consistent with the claims that (1) everything we are responsible for is something that we can in fact control, and (2) control is a sufficient condition for moral responsibility. But since the traditional view holds that moral responsibility depends essentially upon control, my claim is enough to show that the traditional view is mistaken.

In any case, the question of control is notoriously vexing.[13]

3. *Hume and motives*

The view espoused in this paper has its roots in the Humean insight that character is the font of responsibility. Hume's articulation of this insight has been criticized on a variety of grounds. A few comments regarding the differences between Hume's view and my own might help clarify both what my position is and how it addresses some of the objections raised against Hume's view.[14]

[13] See, for example, A. M. Honoré, 'Can and Can't', *Mind*, 1964; Keith Lehrer, 'Preferences, Conditionals and Freedom', in Peter van Inwagen, ed., *Time and Cause: Essays Presented to Richard Taylor*, D. Reidel, 1980; Carl Ginet, 'Might We Have No Choice?', in Keith Lehrer, ed., *Freedom and Determinism*, Random House, 1966; and Peter van Inwagen, 'Ability and Responsibility', *Philosophical Review*, April 1978. In any case, it seems clear that although I am not 'helpless' with respect to my emotions, I do not have any reliable and straightforward method of controlling them. For although I can do various things that might cause my emotional responses (and the beliefs upon which they are based) to change, our present command of psychological principles is sketchy at best, and so the results of any programme of emotional change rather chancy.

[14] Quotations are from David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, Oxford, 1888, reprinted 1968, and David Hume, *An Enquiry Concerning Human Understanding*, in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 2nd edn., Oxford, 1902, reprinted 1955.

Hume asserts that virtue and vice are naught but the causes of pleasurable and unpleasurable moral sentiments. These sentiments arise when an evaluator contemplates the motives that cause an agent to act. Hume concludes that agents are responsible, ultimately, for acts committed at liberty and caused not by passing or casual states of the agent, but by motives that are not themselves considerations of morality, and whose contemplation excites moral pleasure or displeasure.

An action, or sentiment, or character is virtuous or vicious; why? because its view causes a pleasure or uneasiness of a particular kind . . . To have the sense of virtue, is nothing but to feel a satisfaction of a particular kind from the contemplation of a character. The very feeling constitutes our praise or admiration.[15]

'Tis evident, that when we praise any actions, we regard only the motives that produce them, and consider the actions as signs or indications of certain principles in the mind and temper. The external performance has no merit. We must look within to find the moral quality . . . the ultimate object of our praise and approbation is the motive, that produc'd them.[16]

all virtuous actions derive their merit only from virtuous motives, and are consider'd merely as signs of those motives.[17]

no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality.[18]

no action can be laudable or blameable, without some motives or impelling passions, distinct from the sense of morals . . .[19]

Actions are, by their very nature, temporary and perishing; and where they proceed not from some *cause* in the character and disposition of the person who performed them, they can neither redound to his honor, if good; nor infamy, if evil. [A] person is not answerable for [immoral deeds if] they proceeded from nothing in him, that is durable and constant. . . .[20]

Men are not blamed for such actions, as they perform ignorantly and casually, whatever the consequences. Why? but because the principles of these actions are only momentary, and terminate in them alone.[21]

For what is meant by liberty, when applied to voluntary actions? . . . By liberty . . . we can only mean a power of acting or not acting, according to the determinations of the will, that is, if we choose to remain at rest, we may; if we choose to move, we also may.[22]

Liberty, according to that definition above mentioned, in which all men agree, is also essential to morality, and that no human actions, where it is wanting, are susceptible of any moral qualities, or can be the objects either of approbation or dislike . . . it is impossible that [actions] can give rise either to praise or blame,

[15] *Treatise*, III. I. ii (p. 471).
[16] Ibid., III. II. i (p. 477).
[17] Ibid., p. 478.
[18] Ibid., p. 479.
[19] Ibid., p. 483.
[20] *Enquiry*, VIII. II (p. 98).
[21] Ibid.
[22] Ibid. VIII. I (p. 95).

where they proceed not from these principles, but are derived altogether from external violence.[23]

These passages reveal several errors, or so I shall suggest.

Firstly, Hume's claim that liberty, suitably defined, is a precondition for moral responsibility opens the door to a variety of objections. If determinism be true, it may be argued, Hume's view runs afoul of the Principle of Alternate Possibilities and the dictum that ought implies can. Moreover, some[24] have charged that upbringing and heredity constitute 'external violence'. That is, although we can will not to be a miser, we cannot will to desire to will not to be one.

Whether or not such arguments succeed is a rather complicated question. I have argued above that it is preferable to deny that liberty is a precondition for moral responsibility, and that there are better answers to these objections than arguing, as most compatibilists have done, that determinism does not undercut the appropriate senses of 'can', 'could have done', etc. Such attempts have met with much controversy. My view, which does not make liberty a precondition, avoids that particular thicket altogether.

Hume, however, cannot adopt this suggestion. For his claim that responsibility depends ultimately upon *causes* of actions does not enable him to dismiss the causal genesis of an act as irrelevant to moral responsibility.[25]

Here, then, is a second error; Hume is quite wrong that persons are only blameable for x if they had an insalubrious motive that *caused* x. For blameable motives may render one responsible even if those motives were not, in the event, causally efficacious. Indeed, my variant of the Frankfurt counterexample suggests that one can be responsible for something one did not cause at all. Harry's hand struck the face of his landlord without any assistance from Harry himself; that event occurred quite independently of any desire, intention, decision, etc., on the part of poor Harry. True, what made Harry responsible was his intention to perform the slap. But what is significant is not the *causal* relation of the intention to the slap, but what having such an intention shows about Harry's moral attitudes, dispositions, etc. Harry is culpable, I would suggest, because what I would call his 'worldview', that is, the sum total of his attitudes, beliefs, etc., fails to fit the requirements of morality. One must not be misled by the fact that actions and their causal settings are what generally *reveal* attitudes. What is relevant is that one has morally unsavoury attitudes, not whether those attitudes are causally efficacious.

For example, it seems evident that someone who would have abused

[23] *Enquiry*, VIII. II (p. 99).

[24] See, for example, John Hospers, 'What Means This Freedom?' in Sidney Hook, ed., *Determinism and Freedom*, New York University, 1958. See also Thomas Nagel, 'Moral Luck', in his *Mortal Questions*, Cambridge, 1979.

[25] Of course, one could argue that by 'motive' Hume does not mean a cause of behaviour, although his language suggests otherwise. If so, it is not clear what he does mean. If he meant 'attitude or belief revealed by action', I rest content to have clarified Hume's position.

public office had he or she been given the chance, is less laudable, more subject to moral disapprobation, less morally worthy, than is someone otherwise similar who would not have abused public office. The counterfactually corrupt individual is morally accountable for his evil disposition, though, fortunately, that disposition was never actualized. True, we tend to judge such a person less harshly than we do those who actually commit felonies. This is due to a variety of factors. For example, we have limited confidence in our judgments about such counterfactuals, and we may wonder whether actually obtaining high office would have changed the agent's dispositions. Again, it may be morally worse to be corrupt and seek public office than merely to be corrupt, for the former contains an element of deceit and hypocrisy that the latter does not. In any case, we tend to be distressed by harm done, and so more perturbed by culpable harm than by mere culpability as such. Thus the potential felon, if less wicked or blameable than the actual felon, is no less responsible for his felonious disposition.

Because I take the ground of responsibility to be what our actions (etc.) show about us, rather than their causes, my claim that one may be blamed for one's emotions does not commit me to any particular claims about how emotions are caused. I need assert only that having certain emotions *reveals* something about one's attitudes that may be of moral significance. By way of contrast, Hume seems committed to the rather controversial claim that morally evaluable attitudes *cause* the relevant emotions.

Thirdly, Hume is never quite clear about what he takes the bearers of responsibility to be. He generally speaks as if we were responsible for acts (although he does say that in judging acts we 'regard only' motives). For reasons given above, I suggest that we are responsible not for acts as such, but for properties we instantiate.

Fourthly, Hume seems to think that one is not responsible for one's passing fancies, nor for acts caused by passing fancies; only acts that are caused by enduring motives reflect upon us morally. If he thinks this he is mistaken, for our passing fancies and momentary passions do not occur randomly; they indicate something about us as moral agents. That Jones has passing fancies for women may show that Jones sometimes regards women not as persons but as pleasant objects. But that attitude is not, in any ordinary sense, the *cause* of his giving Jane a carnation.

Of course, if Hume means only that we should not *now* be blamed for acts that show *nothing at all* about what we are *now* like, he is quite right. But since what is revealed by actions such as Jones's whimsical floral gift is, often, not properly described as a *cause* of those actions, to say that we may be held accountable for what such acts reveal about us is to deny that causes of actions are what ground responsibility.

Fifthly, Hume insists that the relevant motive must not itself be or include a moral judgment. And Hume is quite right that if an act's being

virtuous is nothing other than its cause exciting moral pleasure, then to say that an act's cause was the agent's regard for that act's virtue would be 'to reason in a circle'. So much the worse, then, for Hume's account of virtue. For it seems evident that the fact that Jones places great importance upon the morality of his acts while Smith does not is relevant to a moral assessment of Jones and Smith. For surely it is a morally salubrious quality of Jones that he is deeply concerned with the moral character of his acts, and a morally unsavoury quality of Smith that he is not.

Sixthly and finally, Hume begins his argument by assuming that a virtue is merely whatever causes a pleasurable moral sentiment. Thus for Hume the question 'under what conditions is A responsible' becomes 'when does A excite in us a feeling of moral pleasure or uneasiness of which he is the object?' As a result, Hume's argument is unconvincing to those who do not accept his 'moral sentiment' view of ethics. My argument makes no such assumption.

4. *Blame and punishment*

At this point one may be tempted to respond as follows. Perhaps it is true that, in my special sense of 'responsible', Charlie is responsible for his emotions. But my special sense is not what we intend when we make the ordinary language claim that Charlie is not responsible for his emotions. Thus my claim, although true, is of little interest.

Now it may well be true that our ordinary use of the word 'responsible' differs from the special sense employed in this paper. But it does not follow that Charlie's being responsible for his emotions, in my sense, has little import. As I have suggested, it is enough to warrant moral evaluation of Charlie on the basis of his emotions. It is even enough, I will argue, to satisfy the 'desert' requirement for punishment. For if we define 'abstract justice' as a match between one's life situation and one's moral situation, then abstract justice demands that Charlie's life situation be inferior to that of someone otherwise just like Charlie who does not enjoy the suffering of animals. And abstract justice is the basis for the desert requirement for punishment.

Abstract Justice. The word 'justice' is used in many ways. In its widest sense, 'justice is served' means that what happens to individuals matches or fits their moral positions. Thus if A is careless and is injured thereby, we might say that A's injuries are 'just', whereas we might bemoan the injustice of an injury incurred after taking stringent precautions. Similarly, we say that it is unjust that a good person is continually ill while a wicked person enjoys uninterrupted good health. Finally, when we say that it is unjust for one felon to receive a five-year sentence and another to receive a suspended sentence, we mean (or might mean) that since the two felons' moral

positions are the same, their life situations should be the same. What all these cases have in common is the insistence that the circumstances of one's life concord with one's moral merits. Let us call this conception of justice 'abstract justice'.

Now it is clear that if, as I have urged, Charlie is responsible for his emotions (in my restricted sense), then the moral quality of his emotions affects his merits, his moral situation. If his emotions are morally heinous emotions, then Charlie is morally inferior, morally flawed. And so abstract justice demands that Charlie's life situation be (to that extent) less than perfect.

Preliminary remarks about punishment. Discussions of punishment have tended to conflate several distinct concepts. It will prove useful to disentangle them before proceeding.

Firstly, moral responsibility should not be confused with (justifiable) criminal liability. If A decides to punch B when next they happen to meet, but fails to do so because A happens to be tied to a chair during their next encounter, A is morally responsible for his malign intention but is not legally liable (he made no attempt, since he took no step, preliminary or otherwise, toward bringing about the assault, and is guilty of no conspiracy, since no others are involved). Conversely, regulatory law provides for instances of non-culpable criminality, though the courts have countenanced strict liability only when the penalty is slight ('not truly criminal').[26]

Secondly, one should not conflate licitness or permissibility of punishment with appropriateness of punishment. For punishing felons is a social policy, and, like all social policies, is subject to two sorts of evaluations: (a) is the policy wise, does it cost-effectively serve valid social goals, and (b) is it permissible? The latter is a question about the legitimacy of the policy, the former about the wisdom and advisability of the policy. Generally, questions of permissibility take priority; they set the limits within which social policies may operate. But unless one is a strict Kantian, one must admit that the permissibility of a practice depends somewhat on its usefulness and its cost. (It is at least arguably permissible to kill one citizen in order to save every other human being on the planet, but it is clearly not permissible to kill twenty citizens in order to save one dog. Thus the cost, necessity, and usefulness of the act influence its permissibility.)

Questions about the permissibility of punishment generally involve two questions: is the felon subject to punishment, and are we entitled to mete it out?

The question of whether x is subject to punishment is either a question of the needs and claims of others (extrinsic subjection) or a question of abstract

[26] Cf. *United States* v. *Dotterweich*, 320 U.S. 277 (1943), *United States* v. *Park* 421 U.S. 658 (1975), etc.

justice (intrinsic subjection). That is, A can be subject to punishment either because he merits it (intrinsic) or because others have entitlements that can only be satisfied by punishing A, and those entitlements are stronger than any entitlement of A's *not* to be punished (extrinsic). Abstract justice is the conception that a person's living situation should accord with his moral situation. So x is intrinsically subject to a worse life if his moral situation is worse than his current life situation.

But even if this condition is satisfied, it does not follow that we are entitled to mete out abstract justice. It may be, for example, that a good man is more entitled to live than a bad man, but, human judgment being fallible, no human being is entitled to kill on the basis of his judgments of relative merit. If so, the bad man is subject to being killed in order to save the good man, but no one is entitled to do the killing. Similarly, it might be that, in some sense, a perennially drunk or reckless driver 'deserves' to be killed or injured in an accident, but not that I or anyone else has the right to arrange for such an accident to occur.[27]

Finally, even if a system of punishment is permissible, it does not follow that it is a good thing to punish felons. And a bad thing that causes suffering is clearly condemnable. Thus in order to punish a felon, A, we need to show that:

(1) either A's moral situation merits a worse life, or someone else has an entitlement requiring A's punishment that overrides A's claims not to be punished, and

(2) We are entitled to be the instrument through which A obtains his just deserts, and

(3) punishing him serves some cost-effective social purpose, either *per se*, or because (a) there is a social policy mandating A's punishment that is *generally* cost-effective, and (b) making an exception to that policy is not cost-effective.[28]

Now it seems clear that, at least in some cases, if an offender is truly guiltless, no one else has a relevant overriding entitlement. That is, I will take it for granted that exempting the truly guiltless has little effect on deterrence, and that, for example, the victim's family's desire for vengeance does not override the guiltless offender's claims to avoid punishment.[29]

Thus condition (1), the desert requirement for punishment, is simply

[27] It is here that disagreements between liberals and non-liberals often surface. Joel Feinberg, unlike the Ayatollah Khomeni, denies that the state is entitled to punish heresy, even if such punishments would be socially useful.

[28] Of course if Kant is right, then punishing felons fulfils a moral obligation of society, and fulfilling our moral obligations is a useful social purpose. Thus condition (3) need not be 'utilitarian' in the narrow sense.

[29] If the offender is dangerous, however, others' claims to safety may require his or her incapacitation. But since such incapacitation is not punishment, the incapacitation should be made as pleasant and beneficial to the offender as available resources permit.

the need to show that the offender is intrinsically subject to punishment (he is sufficiently morally blemished).

I will argue that if A is responsible for x in my restricted sense of responsibility, that is sufficient to satisfy condition (1) (though punishing people for their emotions, for example, does not satisfy condition (3), and so is not a justifiable practice).

Deserving punishment. Charlie, we have seen, is a morally worse person because of his enjoyment of bestial agony; he is a morally imperfect, morally faulty being to the extent that he has such feelings. Thus it seems clear that abstract justice requires that Charlie not have as good a life as Mike, who is similar in every other respect but does not enjoy the sufferings of beasts. For Mike is more morally worthy than Charlie. Thus any scheme that gives each person his moral due, that distributes goods according to moral worth, would give Mike more of the good things of life than Charlie. Conversely, Charlie cannot justly complain that Mike fares better than he does, for Mike is morally superior to Charlie. Thus to the extent that Charlie's feelings are morally odious, he 'merits' punishment. That is, his worsening of life situation is in accordance with a principle of just distribution of life situations. Thus there is an important sense in which Charlie 'deserves' or merits harm as a result of his emotions.

Of course, Charlie may have a just complaint on other grounds. It may be that no one has a right to ensure that Charlie fares worse than Mike (condition (2) might not be satisfied). If so, punishing Charlie is unfair, just as it would be unfair to deny a miscreant his inheritance. Were the miscreant to lose his inheritance, we would feel that justice had been served. But it would be unfair for us to refuse the miscreant his inheritance in the cause of justice. For we have no right to interfere in that way. Similarly, it may be that condition (3) is not satisfied; no useful or worthwhile social purpose is served by punishing Charlie. If so, Charlie might make a strong case that inflicting suffering upon him is unjustifiable. For if suffering, as such, is undesirable, then there must be some good reason for deliberately inducing it.

Thus it does not follow that we *ought* to punish Charlie. My point is merely that the 'desert' or 'minimal retributivist' requirements for punishment are met; there is a moral basis for deeming that Charlie's life situation deserves to be worse than someone otherwise like him but without his heinous emotions. Whether we ought to act to *ensure* that Charlie's life situation is in fact worse in yet another matter.

Conclusion. We are responsible (in my limited sense) for our emotions, even if we did not choose them. That limited sense of 'responsibility' is sufficient to warrant moral evaluation (praise or censure) and to satisfy the desert requirement for punishment. It follows, then, that contrary to prevailing

moral sentiment, deserved praise, censure, and perhaps punishment and reward do not presuppose control. If I am right, many of our moral beliefs need to be rethought.

Department of Philosophy
Swathmore College
Swathmore, PA 19081
U.S.A.

EUGENE SCHLOSSBERGER