

# Morality without Responsibility? How to Be Thoroughly Immoral Without Being Guilty

Kenton Machina

## Introductory remarks

These are excerpts from the notes for a colloquium I gave at ISU in the Fall of 2007. The full colloquium included arguments for the definition of moral responsibility, but that part of the colloquium is omitted here. I have included, however, some clarification of what “moral standing” means.

The main part of the colloquium that appears below deals with the question whether giving up on moral responsibility as a legitimate concept would imply that morality in its entirety should be seen as illegitimate. In other words, is it conceptually possible, without contradiction, to maintain an active morality while abandoning moral responsibility entirely?

One author we will study – Pereboom – argues that it is possible, and in fact would be desirable. He believes no person is ever morally responsible in virtue of anything. (Where MR is defined more or less as we have defined it.) But he definitely wants to preserve a moral system for guiding behavior. So, he needs it to be coherent to have morality without MR.

If it truly is possible to coherently maintain a serious moral system to guide behavior while giving up on MR, then the cost of giving up MR goes way down, and more people might be willing to go that route. But if giving up MR means giving up on morality entirely, then many people would say the cost is too high.

It has been assumed by many that giving up on MR implies that morality itself makes no sense. Pereboom and certain other philosophers have recently questioned that assumption.

In this talk, I argue that only certain versions of morality can survive if there is no MR. So, in a way I am agreeing with Pereboom. However, I am also saying that in a way he has oversimplified the situation, by overlooking the fact that certain very attractive moral ideas require MR. So, if I’m right, the cost of giving up MR is higher than Pereboom indicated, because one would have to give up on certain rather commonly held versions of morality.

I do not argue here for any particular version of morality. In fact, however, I actually do favor a version of morality that cannot survive if MR is given up, so for me the cost of giving up on MR is very high, since my favored type of morality would also have to be given up.

But nothing in this paper depends on which version of morality is best. The type of morality Pereboom himself seems to be happy with does in fact survive if MR is abandoned, as far as I can see. So, Pereboom would not necessarily find this paper objectionable.

Because these are notes from a colloquium, they are not an essay. Sometimes the wording may not be clear. If you find something unclear, just make a note of your question and continue on. We can clear up questions in class.

Parts of this discussion are fairly subtle and detailed. You have to pay close attention to detail. If you get lost, make a note of what you don't understand and we'll try to fix it in class. Don't give up just because the reading gets complicated toward the end. But plan to spend some time with the reading – skimming isn't going to work.

## **The issue**

My dog, Shanu

Sometimes he's a good dog, sometimes he's a bad dog. When he's good, he deserves praise – he's praiseworthy. When he's bad, and gets caught, he becomes the subject of certain reactive attitudes from me, including indignation, and he is blameworthy. He can be trained to be good more often.

I use these terms – praiseworthy, blameworthy, good, bad, indignation – advisedly. I think they are being used in a perfectly normal, literal sense. But these are some of the same terms as appear in serious discussions about moral responsibility. It is often said, for example, that to be morally responsible for having done something morally wrong is to be morally blameworthy.

My dog is not morally responsible, and when he's bad, he's not morally bad; dogs are not capable of being evil, sinful, or morally depraved – he's not *morally* blameworthy. When he's good, he's not *morally* praiseworthy. Dogs are not capable of being saintly, holy, or morally good. Morality applies to persons, not dogs. At least this is how it seems to me, and I assume these claims are uncontroversial.

But now we can ask, Are people just like dogs, at least with respect to these matters?

I take it that one traditional line that rejects MR implies an affirmative answer. People sometimes do things that are socially unacceptable, or socially approved, and people can be trained to a certain extent, all just like dogs. But, again just like dogs, people are not capable of moral responsibility and there are no moral "oughts" since the truth of causal determinism of human behavior plus the axiom that "ought" implies "can" rules out moral "oughts". We physically *can* do only what we actually do. There is no ought to do what we don't do, so there is no possibility of doing moral wrong. The development of scientific understanding of the causal structure of the world thus leads enlightened folk to abandon outmoded ways of trying to control human behavior through morality, and to move instead

to something more sensible, akin to dog training. Morality doesn't apply to people any more than it does to dogs, according to this way of thinking.

However, for some of those (like Pereboom) who think the causal structure of the world precludes moral responsibility, the picture of people as being like dogs in this way is unnecessarily reductive. These folks believe people can coherently be subject to a genuine morality, even without moral responsibility. This view, which Pereboom prefers to term "hard incompatibilism", drives a wedge between MR and morality, and claims you can have one without the other. More carefully put – you can have morality without MR. (Not the other way around – clearly MR without morality makes no sense.)

Can there be morality without MR?

Meaning:

Suppose we consider a culture in which there is individual moral right and wrong, but no moral responsibility. So, even though I live a thoroughly immoral life, and everyone knows it, I am not thought to be morally responsible for any of my misdeeds. And the same is true of everyone else, no matter how morally upright, or immoral they may be. My question: is such a conceptual framework coherent? Some would say obviously not. Some would say that without moral responsibility, there logically can be no morality.

Another way to see the importance of this

If it is possible to have robust morality without MR, the social/conceptual price of giving up on the troubled idea of moral responsibility would not be nearly as great as the price were morality to imply MR. So, related to the central issues surrounding MR and its acceptability – not just Pereboom's argument.

So, can there be morality without MR? Try this thought experiment:

Set up a society of MR rejectors in which there nevertheless is a legal system to determine whether someone broke the rules and to meet out enforced consequences such as imprisonment for misdeeds. Suppose on the more personal level there is talk of what behaviors are right and which are wrong, and that this connects with views about praiseworthiness and blameworthiness, with actual praising and blaming seen as techniques for shaping people's behaviors. This means that to be praiseworthy is simply a matter of having done something that people desire to promote and that has a good chance of being promoted by overt praising. But in our imagined society everyone agrees that the way each member of society behaves is entirely due to the causal structure of the world – all choices made by members of the society are due to the causal influences of previous experiences, previous physical interactions with the world, and genetics. It is agreed by the folks in our society that this means no one is ever truly MR for their own actions. In this regard, at least, people are thought to be like dogs. However, since talk of right and wrong is used in a sophisticated way to influence behavior, these folks believe they have a robust morality, without MR, and so in this way people are not like dogs.

This little story doesn't really settle the matter

- 1) Does this society have a genuine morality? Or is it really just dog training?
- 2) Does this society truly lack MR? Or do they simply refuse to apply the label MR to what really is MR?

The concepts of morality and MR are in need of clarification prior to employing them in arguments.

E.g., one could simply *define* MR as requiring the agent be the sole, uninfluenced author of whatever she is MR for, thereby ensuring that MR is unreal for human beings. Some arguments on this topic may amount to little more than this. ("How can one be responsible for something for which one is not the ultimate cause?")

Too easy.

### **What is the concept of MR?**

MR is a term of art in philosophy, intended to label a potentially fuzzy set of related moral concepts, such as moral guilt, moral praiseworthiness, moral desert, and the like.

What seems to be at stake in being MR is the potential for affecting an agent's "moral standing" – the person's overall level of moral excellence or its opposite. This is the basis for desert of the relevant sort, the basis for legitimating the relevant reactive attitudes.

Try this for a more precise description:

**A person is MR in virtue of some act, omission, or mental state just in case the person's "moral standing" is partially a function of the "moral standing" of that act, omission, or mental state. This is the only way a person can be MR.** (In many cases, the relevant acts, omissions, or mental states are the person's own. There may be extended cases in which these in some sense belong to someone else. E.g., perhaps a parent would be morally responsible for racial bigotry in their six-year-old daughter.)

What I mean by "moral standing" of an act, omission, or mental state: the verdict that an accurate moral evaluation of that act, omission, or mental state would render. Examples:

Morally right, morally wrong, or morally neutral

Morally supererogatory, morally permissible, morally impermissible, morally outrageous or despicable

Unjust, just, just but harsh

What I mean by "moral standing" of a person: the verdict that an accurate moral evaluation of that person as an agent would render.

Examples:

Moral saint, moral jerk, morally bankrupt, morally exemplary, morally acceptable

Decent sort of person

Important point of clarification re the “moral standing” of a person:

There are puzzles about the MR of people who are being manipulated, of children, of people in various dysfunctional mental states, and so on. So, it has to be conceptually possible that a person exhibit a pattern of morally significant behavior for which they are NOT MR. This means in my account the “moral standing of a person” cannot consist of dispositions to behave morally or immorally, and cannot merely consist of patterns of typical behavior.

E.g., “Joe will take advantage of others whenever he can” is not to **automatically** count as part of Joe’s moral standing. It is just a description of Joe’s behavior patterns that leaves open whether he is MR for being that way. It does not say whether our evaluation of Joe’s moral standing should be influenced by this pattern.

Descriptions of morally significant patterns of behavior or thoughts are not descriptions of what I mean by a person’s moral standing.

Descriptions of a person’s psychology or dispositional states are not descriptions of what I mean by a person’s moral standing.

In daily practice, we often conflate the evaluation of the person with the evaluation of their behavior patterns or character traits because we tacitly assume the person is MR for being the way they are. So, in practice we condemn or praise people by describing their behavior patterns. But for our purposes here it is important to bring such tacit assumptions out into the open. For our purposes, it is important to distinguish between the evaluation of morally significant behavior patterns and the moral evaluation of the person which MR implies.

Again, the moral standing of a person is the verdict that an accurate moral evaluation **of that person** would render – it is *not* a description of morally relevant features of that person’s behavior patterns or mental states. **It is an accurate evaluation of the person’s moral goodness or badness** in virtue of behavior, omissions, mental states, and anything else that may be relevant.

Whether the moral standing of a person implies something about what they deserve or about the cleanliness of their soul is not officially part of the account. Nothing in my account rules out an attempt to draw such connections.

Whether the moral standing of a person can ultimately be connected to how the person ought to be treated is also left out. I don’t see any barrier in my approach to building such an account.

It will become tiresome to keep saying “actions, omissions, or mental states”. Morally significant omissions are something like full-blown actions in that they are in a sense voluntary. And those who believe mental states can be morally assessed are treating assessable states as being like voluntary actions. So, I’ll just talk about actions from now on, and mean that to include omissions and mental states as appropriate in context.

Given the core account of MR above, it is clear that MR implies morality. So, the next question is whether morality implies MR.

### **The main question: morality without MR?**

To have morality without MR means

Some acts are morally right, some are morally wrong, but no one has moral standing as a morally good or morally bad person.

You might deliberately and knowingly engage in the worst sort of behavior without ever taking on the moral status of being a morally bad person. That means it is inappropriate, if we use Strawson's account of blame, for others to harbor moral resentment or other negative moral reactive attitudes toward you. Or, in terms of a desert theory of MR, you are not deserving of blame.

The issue is whether this sort of thing is logically coherent. Probably the majority view is that it is not possible. Traditionally, it seems most have assumed that morality requires MR, so that giving up on MR means abandoning morality. But some (like Pereboom) who are suspicious of MR, but don't want to give up on morality, have assumed that it is possible to keep morality and give up on MR.

At this point the work done above on the analysis of MR becomes crucial. Using that analysis of MR, here is our question: can acts routinely be morally right or morally wrong without any person ever having moral standing in virtue of those acts?

On the traditional view,

when there are not excusing conditions present, the agent's moral standing is a function of the morality of the agent's actions, and excusing conditions cannot be present most of the time, so morality requires MR.

However, morality applies to **actions**, while MR is a feature of **agents**, and that may open up enough of a wedge to challenge the traditional view.

Perhaps we are not forced to assess **persons** just because we assess something those persons do. **Think of a child rearing strategy - "That's naughty and you shouldn't do it, but the fact that you did it doesn't mean you should think you're a bad person." Why can't we apply this sort of thinking across the board?** It takes morality seriously as a guide to conduct, but never assigns any moral standing to anyone.

This seems to be Derk Pereboom, Bruce Waller, Michael Slote (to some extent).

(I am not interested here in whether this strategy would be desirable, but only whether it is coherent.)

**The key claim in this paper: the wedge strategy works only for some versions of morality.** That is, we can coherently maintain morality without MR only if we adopt a certain sort of morality – what I shall call an “externalist” morality. But if our morality is a more internalist, “motivational” morality, we will be required to maintain MR as well. (This is different conclusion from the one in my *Acta Analytica* paper that some have seen.)

#### “Motivational” morality

The moral evaluation of an act depends at least in part on the motivational structure of relevant agents – motivational structure that accounts for or explains what actions are being performed. The morally relevant aspects of that structure are to be built into the action description prior to its evaluation, and the evaluation is applied to the result.

E.g., if I donate toys for a Christmas toy drive, believing that I am thereby poisoning large numbers of children and hoping to thereby destroy their lives, I am doing something far more than voluntarily donating toys. I am attempting to poison innocent children, which makes my activity morally wrong even if my belief is mistaken and my donation brings about nothing but good results.

One way of reading Kant makes his moral theory count as motivationalist – the morality of an act token depends on the universalizability of the maxim with which it is performed on that occasion, where the maxim is determined by the agent’s motivations.

But what about typical cases of immoral negligence in which the agent is not intending to be negligent or careless? The motivational approach does not look at *intentions* alone, but also at all other morally relevant components of the agent’s motivational structure. Negligence, then, might be thought of as not trying hard enough, where the degree to which one tries to be informed, to do the right thing, counts as part of the motivational structure, thus serving as a basis for the evaluation of the agent’s activity.

#### “Externalist” morality

The moral status of an act does not depend on the motivational structure of the agent, but rather on other factors such as the consequences of the act, or whether the act is in accord with natural law, God’s laws, or a universalizable moral principle that does not invoke reference to intentions. Classic act utilitarianism is an example.

E.g., my toy donation is morally permissible, and perhaps even laudatory, despite my intentions, because donating many good, safe, new toys to a charity when one can afford to do so is a morally laudatory thing to do.

If we are operating from within an externalist morality, it seems quite possible to have morality without MR. The morality or immorality of an act does not imply anything about the moral standing of the agent.

To see why, consider the cases in which it seems the most difficult to separate morality from MR: *life-long patterns* of immoral behavior. On an externalist view, these imply something about character, where character is taken to mean simply what patterns of moral or immoral behavior we can decipher within the agent's history. Recall that this sort of character-as-behavior-patterns is NOT the same thing as moral standing. One might acknowledge the pattern of moral misdeeds while refusing to draw any conclusions about the agent's moral standing, by claiming that the entire pattern was not ultimately up to the agent, perhaps because of social influences or parental failures.

Agents on this view might be considered dangerous even if not morally blameworthy, or be thought to be very good at helping everyone around them to flourish without being considered praiseworthy. There is nothing in the externalist point of view that *requires* that we attribute any moral standing to any agent in virtue of the moral evaluation of acts. To evaluate an act is not necessarily to evaluate any agent's moral standing, even though historically it may well be that most externalists have indeed assigned moral standing to agents on the basis of the moral evaluation of their actions. So, now you know how to be a very bad person who is not at all guilty (in the appropriate sense).

In contrast, from the point of view of an "motivational" morality, I will now argue, it is not possible for a society to maintain morality without including MR in typical situations.

The argument for this claim comes in two versions, depending on how the motivational morality works:

Version I – the easier case: the moral assessment of an act includes moral assessment of the relevant motivational structure from which the act flows.

Example: my toy donation is wrong because my motives are wrong.

Version II – the harder case: moral assessment of acts must always take into account the relevant motivational structure lying behind acts, but does *not* assess those motivations per se. Instead, it is the act, described at least partly in terms of its motivation, that is assessed.

Example: the toy donation, not the motives behind it, was wrong, because the donation was actually an attempt to harm innocent others. On this version of motivational morality, it is the donation, not the motivation, that is assessed, but the motivation gets built into the description of the act, resulting in the act's being described as an attempted poisoning. It is the act so described that is morally evaluated.

Version I argument – the motivational structure associated with an act is evaluated as moral or immoral:

When one assesses an agent's motivational structure as morally acceptable or unacceptable, one is necessarily assessing that agent as morally good or bad on that occasion. For the motivational structure cannot be separated from the agent – it is too closely tied to the agent. So this type of morality assesses an agent's moral standing in virtue of an act. As argued earlier, that amounts to MR.

There is no way to say “Your motives are all immoral, but that says nothing about your moral standing as a bad person on this occasion”. One might even hold the view that the motivational structure of an agent is metaphysically a central part of what makes up the agent.

Version II argument:

Now we consider moralities that do not normally evaluate motivations, but rather evaluate actions considered in terms of their motivations.

Examples of such descriptions:

trying to poison an enemy

attempting to provide emotional support

getting even, seeking revenge

providing shelter for a lost animal, out of sympathy based on childhood experiences

My argument regarding a morality that focuses on such motivation-laden descriptions of the actions up for evaluation:

The basic claim: these act descriptions involve “enough of” the agent that there is usually no way to separate the agent’s moral standing from the evaluation of the act. Even though the agent’s motivations are not isolated for evaluation, the motivations are built into the complex that is evaluated. Consequently, the evaluation of the act complex normally indicates an evaluation of important aspects of the agent. The agent is thus seen as someone whose moral standing is potentially affected by the evaluation of the act complex. That is, the agent is MR.

Examples:

If I am trying to poison large numbers of children, and I am of reasonably sound mind, and not under extreme extenuating circumstances, then to pronounce my act as morally abhorrent is to say that I am on this occasion a morally bad person, not just that I am doing something morally wrong that might not reflect on my moral standing. Or, to reverse the case, if I am trying to be a generous and supportive citizen, with love in my heart for poor children, and I donate large numbers of toys to them, not realizing they are all poisonous, and not negligent in my ignorance, then the moral evaluation of my act, using a reasonable version of the motivational approach, will normally include something to the effect that I am a morally good person at that time, despite the miserable consequences. (Whether it evaluates the act complex as good or not doesn’t matter for the argument.)

There is an objection, though. The objector points out that in many moralities, some circumstances arise in which morally significant actions do not reflect on an agent’s moral standing. E.g., someone does something wrong, but is excused because of unavoidable confusion, and so there is no effect on her moral standing. The objector, seeking to defend the possibility of morality without MR, asks why

one could not have a *motivational* morality in which everyone falls under some such circumstances all the time, so that morality remains but MR disappears. (Maybe the generic circumstance that eliminates MR would be based on the truth of determinism.)

My response to the objection:

Let's focus on intentional action, because MR may be the most difficult to eliminate from motivational morality in the case of such actions, for reasons already given – intentions seem to be part of the agent. My aim is to show that the objector cannot maintain moral evaluation of intentional action while simultaneously abandoning MR. If I can do that, the objection has been dealt with sufficiently.

Suppose I intentionally cause you considerable pain even though I believe you don't deserve it; I just get pleasure out of hurting others. The objector has to propose some reason that this motivational structure does not reflect on my moral standing; after all, the intuitive response of most folks will be that it does reflect quite negatively on my moral standing. The only reasonable route open to the objector is to claim that the intentions are not truly my own, even though I do have them. I have to be separated from my intentions in this case. The only way to do this is to claim that crucial facts about the motivational structure alienate me from these intentions – perhaps something to the effect that I am the mean sort of person that I am because of my personal history or my genetics, and that there is nothing I could do about it, so that my intentions are not owned by me, since I am not their source. Or maybe my intentions are caused by something in the air, or by a random intention generating machine somewhere. They are said thus to be alien to who I am. And since the objector's aim is to eliminate MR in general and not just my own MR, the claim will have to be extended – the intentions in me don't reflect on any other agent's moral standing either. (My intentions haven't been inserted into me by someone else.) So the intentions are not only alien to me, but also to everyone else.

Thus, according to the objector, the action complex to be morally evaluated must include the information about the alien character of my intentions, since it is highly morally relevant, and Type 2 motivational morality evaluates action complexes that include all the morally significant aspects of its motivational structure. And, according to the objector, *when this information is included in the action complex, the complex may be morally evaluated, but my intentions no longer reflect on my moral standing, or on the moral standing of anyone else.*

Moreover, this same strategy will have to be employed by the objector with respect to *all* of the morally significant intentions I have. None of them reflect

on who I am, morally, or on anyone else. Thus, the objector asks the Type 2 motivational morality to evaluate intentional action complexes that never include intentions that are any agent's own intentions for which the agent is responsible.

An example might help. Try this: imagine there is a psycho-active drug that not only dramatically relieves depression, but also causes a few people genetically similar to me to hate everyone around them with a burning passion. Suppose this hate effect is unknown to the medical community, and that the drug has been used successfully on many others to treat depression. Then I am prescribed the drug. The hate it creates immediately produces a serious, effective intention on my part to harm you, and the result is an intentional harming action on my part. I don't understand why I hate you so much, but the hate is quite real. I am perplexed, but I do form intentions to harm you and I act on them. I don't understand myself. Suppose the objector would point to this example as a legitimate case of my not owning my own intentions, and thus escaping from being MR for my harmful behavior on a reasonable version of motivational morality.

The objector is claiming that conceivably all of everyone's morally significant intentions are like that (especially if determinism is true), and that this shows that even motivational morality might allow for the possibility that moral assessment of action could continue without MR.

My response:

There is no reason to agree to the claim that the resulting intentional action complexes are morally assessable at all, from the point of view of motivational morality. One who takes motivational structures seriously in the assessment of action complexes would probably say that when I harmed you in the example above, I did not do anything morally wrong, or morally right. The action complex looks more like an accident, or a case of hypnosis on steroids gone awry.

It goes against the grain of motivational morality to look only at the undeserved harm done to you and then to proclaim that a great moral wrong has occurred. If that great undeserved harm to you was done with the best of intentions and no negligence, motivational morality might pronounce the act as morally good, rather than morally bad. So, when the harm flows out of an intentional structure that is not owned by the agent, motivational morality is confronted with an aberrant motivational structure in which moral agency is in doubt. If the aberrant intentions in me had been deliberately implanted by someone else, the resulting action complex might be evaluated as being the act of that other party, using their motivational structure as a key component of the action complex. But when there is no other party involved, as in the case

of the above example, there seems to be no standard motivational account of the harmful behavior. Without a motivational account of the behavior, the motivational moral perspective seems required to deny that the behavior can be morally assessed at all, since it is missing an essential ingredient – understandable motivation.

So, the objector has not provided us with a good reason to believe that MR can be avoided while motivational moral assessment of intentional action continues.

This contrasts with the way all this works out from the externalist point of view. In an externalist morality, one may morally evaluate intentional actions performed by agents who are capable of being motivated by moral considerations, without looking at the motivational structures that produced the effective intentions. If I harmed you intentionally, and you were innocent, I did wrong, even though I may escape from MR because I was acting under the influence of the drug. This allows for a moral evaluation of the action to be sufficiently distinct from the moral evaluation of the agent to make morality without MR conceptually possible. This is so regardless of whether the externalist accounts for the moral wrong in terms of rights violation, in terms of causal consequence, in terms of violation of natural law, or in other externalist terms.

Motivational morality has some considerable intuitive appeal. As I read Kant, it has his backing. It got the backing of the ND Supreme Court in *State v Leidholm* (1983). So, if I'm right above, it is not correct to say that on just about any plausible account of morality, one can maintain robust morality while jettisoning MR.

Not surprising that the viability of morality without MR depends on the way morality is structured. MR is a moral notion. Questions about its viability are necessarily at least to some large extent moral ones.

Kenton Machina  
November 2007