

Journal of Cognition and Neuroethics

The Limits of a Pragmatic Justification of Praise and Blame

Ryan Lake
Clemson University

Biography

Ryan Lake earned his PhD in Philosophy from the University of Miami in 2013. His main research interests are in ethics and metaphysics, in particular the problem of free will and moral responsibility, and related issues regarding social policies and criminal justice. He is currently teaching as a full-time Lecturer at Clemson University.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). March, 2015. Volume 3, Issue 1.

Citation

Lake, Ryan. 2015. "The Limits of a Pragmatic Justification of Praise and Blame." *Journal of Cognition and Neuroethics* 3 (1): 229–249.

The Limits of a Pragmatic Justification of Praise and Blame

Ryan Lake

Abstract

In recent decades, many philosophers working on the free will problem have been attracted to a kind of approach, developed by P.F. Strawson, that justifies belief in free will and moral responsibility by appeal to the essential roles that it plays in our personal and social lives. In this paper I explore some of the limits of this sort of pragmatic approach, arguing that while it may provide a strong justification for treating people as free and responsible in some contexts, especially in our personal relationships, there are reasons to think that this kind of approach is not enough to justify our harshest retributive impulses, especially in contexts like that of a criminal justice system.

Keywords

Free will, moral responsibility, compatibilism, incompatibilism, criminal justice

In recent decades, a kind of pragmatic approach to questions of free will and moral responsibility has gained popularity. The popularity of this approach can, at least in large part, be attributed to transformative work of P.F. Strawson (see Strawson 1974). Strawson, and many who follow in his footsteps, argue that belief in free will—in particular the sense of free will needed to ground moral responsibility and the practices connected to it—is justified by the essential role it plays in our personal and social lives. Like many others, I find this kind of pragmatic approach very appealing—but only to a point. In this paper, I would like to explore some of the limits of this approach. I will argue that the Strawsonian framework can provide strong justification for holding people responsible in some contexts, especially in our personal relationships, but that the pragmatic considerations invoked by Strawsonians are not enough to justify our harshest retributive impulses, especially in contexts like that of a criminal justice system.

I. The Strawsonian Framework

First, I would like to very briefly sketch what I take to be the Strawsonian view.¹ I won't have room to defend it at great length in this paper, but I would at least like to say

1. Of course, people interpret Strawson differently, and take different lessons from his work. What I will sketch here is not meant to be a definitive exposition of Strawson's own view. Rather, I aim to sketch some

a bit about why I think many (including myself) find it to be important and compelling. It will undoubtedly strike many readers as strange, if not entirely misguided, to try to ground the existence of free will and moral responsibility in any kind of pragmatic considerations. Many skeptics argue, plausibly enough, that the fact that the belief in free will and moral responsibility is so central to our personal and social lives tells us nothing about whether that belief is true. It would be similar to arguing, for example, the claim that belief in God is necessary for a meaningful and fulfilling life would, even if true, be no epistemic justification for believing in God (though it might be some pragmatic justification).

In response, I would argue (in line with Strawson, and many others) that claims about freedom and responsibility are fundamentally normative. There is no metaphysical feature of the world we can point at to demonstrate the appropriateness of blame, or to show that a particular agent at a particular time is deserving of praise or gratitude. Claims like these are not existence claims, like claims about the reality or non-reality of God. Rather, they are, at their core, claims about how we ought to regard and treat both others and ourselves, about which kinds of emotions and which kinds of social practices are deserved or fitting or appropriate, and which are not. When making normative claims of this sort, as opposed to simple existence claims, pragmatic considerations regarding the nature and quality of our lives, our relationships, our self-esteem, etc.—claims that are intimately connected with regarding others and ourselves as morally responsible agents—become relevant to the truth of those claims.

To say *is not* to deny the relevance of metaphysical considerations to claims about freedom and responsibility. In my view Strawson and some others are mistaken to conclude that metaphysical considerations are completely irrelevant to claims about freedom and moral responsibility. This is because some metaphysical considerations are *included* in the normative standards involved in evaluating the appropriateness of praise and blame. As Gary Watson famously argued, for example, when we learn enough detail about precisely how someone came to be the kind of person they are, even if the person in question is someone truly monstrous (as in his Robert Harris example), our intuitive judgments of freedom and responsibility can be substantially altered (Watson 2004). Drawing on considerations like these, a number of philosophers, working well within the Strawsonian framework, have developed strong arguments for skepticism about free will and moral responsibility.

broad lessons that I, and I think many others, draw from taking the sort of approach to the problem of free will and moral responsibility that Strawson did.

In what follows, I will very briefly sketch a few different sorts of skeptical worries that challenge belief in free will and moral responsibility. My aim in this paper is not to evaluate whether or not any of these arguments ultimately succeed. I only wish to show that the following skeptical arguments are at least *prima facie* plausible, but not (at least given the current state of the dialectic regarding free will and moral responsibility) decisive. The central question I want to consider is how viable a pragmatic justification for belief in free will and moral responsibility (and the practices connected to this belief) is in light of such worries.

II. The Standard Incompatibilist Arguments

A main source of skeptical worry is, of course, the standard arguments for incompatibilism. In recent decades in particular, some powerful new incompatibilist arguments have been developed and much discussed. There is Peter Van Inwagen's familiar Consequence Argument, considered by many to be among the strongest arguments for incompatibilism. It is given various formulations—here is a relatively informal one:

If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (Van Inwagen 1983, 56)

It is easy to see the force of this argument. The fixity of the past seems beyond question, and it seems plausible to say that event which is a necessary consequence of something unchangeable would itself be unchangeable. This argument motivates what has been termed a 'leeway' condition for free will—that we act freely and responsibly in a given situation only if some other action was possible.

Other standard incompatibilist arguments work to motivate what have been termed 'source' conditions for free will—the idea that we act freely and responsibly only if we are in some deep sense the *ultimate source* of our actions. Versions of this view have been developed and advocated by a number of different philosophers, notably Galen Strawson (Strawson 1994) and Robert Kane (Kane 1996). This intuition has been cleverly defended by the use of manipulation arguments, developed most notably by Derk Pereboom with his famous four-case argument (Pereboom 2001).

Now of course, many incompatibilists who accept arguments like these are not skeptics—they are libertarians who believe that we (at least sometimes) act with fully

free will. Nonetheless, if either these kinds of classical incompatibilist arguments succeed (or any other incompatibilist arguments), that **increases the probability** that we lack the kind of free will that could ground true moral responsibility, that could legitimize praising and blaming people for their actions. If incompatibilism is right, then we have to be able to rule out causal determinism—as well as demonstrate that we have the right kind of indeterministic control for robust leeway and genuine self-creation—before we can know whether or not we have free will.

And indeed, many libertarians explicitly acknowledge that we have little or no epistemic justification for such beliefs. As Richard Double notes in his excellent paper on the ‘hard-heartedness’ of libertarians, most libertarian thinkers—Roderick Chisholm, Richard Taylor, Peter van Inwagen, Robert Kane etc.—provide very little in the way of any kind of positive evidence that we are the kind of uncaused, self-created entities that satisfy the metaphysically robust conditions for free will and moral responsibility that their accounts demand (Double 2002). As Double discusses, many are quite explicit in admitting that we have no evidence for such claims, notably Immanuel Kant and William James.

III. The Difficulty of Moral Growth

Now I want to briefly discuss a different source of skeptical worry about free will and moral responsibility. The source of worry is based in primarily psychological considerations, rather than philosophical ones.² Let us start with a modest philosophical assumption, assumed by almost all who discuss the free will problem. The assumption is that a necessary condition for the kind of freedom that grounds moral responsibility is that we be able to exercise some degree of control over the development of our moral characters over time. The assumption of a capacity for moral self-cultivation is most explicit in a number of libertarian accounts, as described above. But a number of compatibilists, especially in recent years, have articulated the idea that a condition on freedom and responsibility is some sense of self-cultivation.

For example, Al Mele diagnoses a number of variants of Pereboom’s “four case argument,” saying:

2. For this point, and for much of the discussion in this section, I am heavily indebted to Michael Slote. Slote develops an extensive argument for moral responsibility skepticism along these lines in an unpublished manuscript for a new book on free will, which he was generous enough to share with me.

In each case in this series, Plum played no role at all in shaping his procedure for weighing reasons (say, through trial and error over the years he has been in the business of deliberating). Unlike normal agents, Plum had no control throughout history as an agent over this important aspect of his deliberative style. (Mele 2005, 78)

In this, Mele suggests that normal agents—agents who are responsible for their actions—shape important aspects of their characters, such as the ways in which they weigh moral reasons in deliberation, over time. This is meant to be a compatibilist reading of “control”—Mele is not assuming any radical contra-causal ability to change one’s character of the sort a libertarian might insist on. In Mele’s view, what is important is that the development of one’s character is influenced and shaped by (does not “bypass”) the exercise of one’s deliberative capacities over time.

It is easy to see why even compatibilists would be inclined to develop an account of responsibility that requires an ability to shape the development of our own moral characters. The kind of character one has, after all, determines the kinds of actions one commits. And so if one’s character is *not* within one’s control (even in a minimal compatibilist sense), then it would seem that one’s actions would also be (to whatever degree actions are driven by character) outside of one’s control. As Michael McKenna puts the point, “what is so important about an agent’s having a history that lacks the acquisition of pertinent values through means bypassing her ability to critically assess them is that she thereby *has* a history that afforded her an opportunity to shape her moral personality for herself” (McKenna 2012, 169).

The question is whether this is in fact a psychologically realistic claim about human agency. How flexible are our characters really? To what extent do we really shape and cultivate our own moral personalities over time? And to the extent that we actually can improve our characters over time, to what extent is this really driven by internal processes, or to what extent does it depend on outside help?

One source of pessimism about these questions is research that shows that character traits measured in very young children can have considerable predictive power regarding how they turn out later in life. A famous example of this is the “Stanford marshmallow experiment,” a series of delayed gratification studies led by Walter Mischel (Mischel et al 1972). In the studies, young children (ranging from about 3.5 years to 5 years 8 months) were given a choice. They could either take one treat now (like a marshmallow), or they could wait until the researcher came back into the room, in which case they would get two treats. Some children had the self-restraint to delay gratification and wait until the

researcher returned (around 15 minutes), but others could not and would opt for the immediate lesser treat. A series of follow up studies showed that the young children who were better at delaying gratification performed better on a number of measures—they were judged more competent by their parents, they scored higher on SAT exams, performed better on cognitive tests. The original participants even showed more activity in the prefrontal cortex in brain scans conducted in middle life.

Results like this are striking; few would expect that a 4 year old child's capacity to resist the impulse to immediately eat a treat could have such predictive power regarding a person's success in later life. This capacity—now called 'executive function'—is an important part of one's character, and these studies seem to suggest that it is fixed quite early in life. However, this conclusion has recently been challenged. There is now emerging evidence that the capacity for executive function is more malleable than Mischel's initial research suggested, and can be enhanced with the right sort of intervention and training, especially if the intervention is done in early childhood (Zelazo and Carlson 2012). This is certainly good news, but even if right, it doesn't provide much support for the idea that we are in control of the development of our own characters. When it comes to executive function, at least, it seems that making any sort of improvement over the initial capacity we have early in life depends on substantial outside intervention.

And the same may be the case for other, even more clearly morally significant, character traits. Michael Slote discusses the role that empathy plays in moral education (see for example Slote 2010, especially the first chapter), building on the work of the psychologist Martin Hoffman (Hoffman 2001). Hoffman develops the idea that instilling genuinely altruistic, moral motivation and behavior in children requires a process he calls 'inductive discipline,' sometimes simply referred to as induction. Unlike "power-asserting" strategies of moral education (which involve threats and punishments), induction builds from a child's natural initial capacity for empathy. In induction, parents (or other educators) notice when a child has hurt another and then, in a firm but non-threatening way, direct the child's attention to the harm he or she has caused, getting the child to focus on and feel how things must feel for the one that the child has hurt. This leads the child to *feel* the badness of what he or she has done, a painful emotional experience that is a kind of rudimentary guilt. Hoffman argues that if this technique is applied consistently over time, the child will develop an association between these bad feelings and situations in which harm could be (but is not yet) done, without any intervention from parents or other figures, and this will help motivate altruistic and moral behavior.

The key thing to notice about this model is the extent to which the cultivation of moral motivations and moral behavior, via the cultivation of empathy, depends

on parental intervention.³ This is not *self*-cultivation of moral character; instead, this suggests that the cultivation of moral character depends heavily on others. Without crucial guidance, the early development of one's empathy and one's moral character will be stunted, and later development will be extremely difficult.

Of course, this kind of model is controversial. On some more rationalist-leaning views, empathy is regarded as unnecessary for moral motivation or understanding or growth. I don't have room to say much about that dispute here, but I will note that the claim that empathy is at least a psychologically necessary component of moral understanding and moral motivation for people (even if not logically necessary—maybe some other sorts of possible creatures could grasp and be motivated by morality without it) is at minimum very plausible, and seems well supported by a good amount of psychological evidence. It is well established that individuals who possess little empathy or lack it entirely (in particular associative empathy—the ability and tendency to feel what others feel) have difficulty with both moral understanding (for example, they have trouble drawing a distinction between arbitrary “conventions” and “morality”) and moral motivation. Insofar as this model is plausible, our confidence in the idea that we in any substantial sense craft or shape our own moral characters should be lessened.

I think it should be said that even if all of this is right, we still might be able to exercise *some* level of control over our characters. As Neil Levy notes, we may still exercise a kind of indirect control over our characters—we can attempt to engage in long-term projects aimed at altering our characters (Levy 2002). For example, a person with anger management issues might take classes to learn how to meditate in an effort to become calmer and more amiable in his interactions with friends and family. Or a man with a prejudice against a particular ethnicity might embark on a project of studying the history and literature of the group he is biased against to cultivate deeper understanding and empathy with the aim of overcoming his prejudice.⁴ Or a woman might buy an app like “HabitRPG” to channel her love of video games into the cultivation of good work habits.⁵ And so on.

3. Again, this is a point for which I am indebted to Slote.

4. Slote discusses an example somewhat like this in his unpublished manuscript.

5. HabitRPG is an app that allows people to play a sort of video game, in which they gain familiar rewards (gold, experience, levels, etc.) and risk consequences (loss of health, lives, levels) based on successes or failures at pursuing real life goals. For those of us who have cultivated video game addictions, it can be a highly motivating system for cultivating new habits and behaviors.

There are a few things that are important to note about this kind of indirect ability to shape one's own character. As Levy reminds us, one's character is one's way of seeing the world. The impetus to try to make changes to one's way of seeing the world will seldom happen without substantial outside influence. Further, in using techniques like taking classes, or getting therapy, or even using an app, we are relying heavily on assistance from others to shape our characters. Finally, even with a great deal of help from others, long-term efforts to change or improve character are often met with failure, or only partial success. Changing one's character, with or without help, is extremely **hard**.

All of these points strongly suggest that the control we have over our own character is very limited. It might be a substantial enough kind of control to warrant some kinds of praise and blame in some contexts, especially at the level of personal relationships. But it is enough to ground our most extreme negative reactive attitudes, the kind of wrath or hatred that might drive violence? Is it enough to justify a heavily punitive criminal justice system like ours? This is where things start to seem more dubious, or so I will argue.

IV. In (partial) defense of Strawson

At this point I would like to return to the question of the extent to which a pragmatic approach to questions of free will, and in particular moral responsibility, can be justified. As many have argued, our general view of ourselves and of others as morally responsible agents is deeply connected to our relationships with others and our conceptions of ourselves. Freedom and moral responsibility are essential to the possibility of attitudes like love, admiration, and respect, both for others and ourselves. To abandon the concepts of freedom and responsibility is to severely diminish our emotional and moral lives in many ways. As I suggested earlier, given that the normative role played by claims about moral responsibility, pragmatic considerations such as these are essential to deciding whether they are legitimate. Now I would like to say a bit more about what I think about these pragmatic considerations—just what, exactly, is lost if claims about moral responsibility are not legitimate? And what isn't?

To start, let me say a bit about what I think is **not** lost. Some have argued (see for example Peter van Inwagen 1983) that morality itself collapses without free will or moral responsibility, that without praise and blame there can be no legitimate talk of moral obligations, nor even of right and wrong. One way to reach this extreme conclusion is to start with the idea that determinism means that it is impossible for us to do otherwise than we actually do, and then to argue, in Kantian fashion, that this would mean that there can be no such thing as moral obligation (nor of praise and blame). From there

one can argue (as Ishtiyaque Haji does) that there is no such thing as moral rightness or wrongness; “S has a moral obligation to perform [not to perform] A if and only if it is morally wrong for S not to perform [to perform] A” (Haji 1999, 183).

I want to suggest that perhaps the concerns raised by these sorts of arguments are exaggerated. I agree with those who have argued that there are many commonplace examples in which people have moral obligations that they are unable to fulfill. To use an example of Bruce Waller’s—if I borrow a large sum of money from a friend, and then hit financial hardship and am unable to repay the loan, it is not as if I am suddenly relieved of my moral obligation to repay. Rather, it seems more natural to simply say that I am now stuck with an obligation that I cannot fulfill. Or consider an example from antiquity.⁶ In the Greek tragedy *Antigone*, the title character finds herself with both an obligation to bury her brother and an obligation to follow the king’s law, which prohibits the burial. As Waller notes, “To the Greeks, this seemed an unfortunate situation, but certainly not impossible” (Waller 2011). Haji and some others *do* claim it is impossible, but it is not obvious why this should be so. The claim that we can sometimes have conflicting moral obligations seems to be at least as intuitively plausible as the claim that ought always implies can in every instance. This discussion is a rather quick sketch for the sake of brevity; the minimal point I want to make here is that one can still plausibly maintain belief in moral obligations, and moral rights and wrongs, even if we abandon talk of praise and blame. This point will be important for what I have to say in the next section.

Now I would like to turn to what we might plausibly think **would** be lost if we were to abandon moral responsibility. I think that a number of important moral attitudes would be lost, or at least significantly diminished. For instance, I don’t think there can be sincere regret or apology in the absence of moral responsibility (contrary to what some skeptics, like Waller and Pereboom, argue). Waller and Pereboom are right to say that one can *lament* that one is the cause of harm to another in the absence of moral responsibility, or one can lament that one has failed to live up to one’s moral obligations. But true **regret** and true **apology** essentially involves taking or accepting responsibility for one’s failings. Kathleen Gill puts the point nicely when she argues that an apology without an acceptance of moral responsibility is like saying “I’m sorry” when hearing that a neighbor has leukemia—a mere expression of compassion or sympathy rather than a true apology (Gill 2000). Such expressions of compassion and sympathy are certainly nice, and they definitely have their place, but if all of our apologies were reduced to this,

6. This example is discussed by Bruce Waller (2011), Joseph Margolis (2000), and many others.

then it seems that an essential component of our relationships with others would be missing.

I think this point about regret and apology can be bolstered if we consider a kind of argument common to moral responsibility skeptics. Skeptics commonly argue that there is no morally significant difference between a person in a causally determined universe and a person who is thoroughly manipulated (by say, an evil neuroscientist). This is what Pereboom attempts to show with his four-case argument. And this generally seems to be the view of incompatibilists about moral responsibility and causal determinism. As Waller puts the point, “why should the shaping by fortuitous contingencies not undercut freedom if the same shaping by planned contingencies does?” (Waller 2011, 64). Similar points are frequently made by libertarian incompatibilists. For an example of this, see Robert Kane’s discussion of B.F Skinner’s “Walden Two” story (Kane 1996, Chapter 2).

So let’s grant for the sake of argument that incompatibilists like Waller and Pereboom and Kane are right, that there is no morally significant difference between a causally determined agent and one who has been manipulated by an outside agent. And then let us ask—to what extent could a manipulated agent truly regret her actions? To make the question more concrete, let’s consider a specific example. Imagine a woman named Riley who is being completely manipulated and controlled by a wicked neuroscientist who has planted a device in her head. One day Riley sees a child drowning. She has an impulse to save the child, but that impulse is quickly erased by the neuroscientist, who replaces it with an irresistible desire to turn and walk away instead.

Now suppose that after walking away from the beach and knowingly allows the child to drown, Riley then later learns that her actions had been directly programmed and controlled by a nefarious neuroscientist. It seems that Riley would be right to believe that she was not blameworthy for letting the child drown. Could she at the same time sincerely regret her action? It seems clear to me that she could not. Riley might be extremely sad that the child had drowned, and she might lament the fact that she had been used as a tool by the neuroscientist to bring about the child’s death. But insofar as she *truly* regards the neuroscientist’s manipulation as completely undercutting her moral responsibility, it is hard to see how she could genuinely regret the action. If this is right—and if incompatibilists are right that there is no morally significant difference between manipulation and ordinary causal determinism—then it is also hard to see how a moral responsibility skeptic can say that it would *ever* be appropriate to experience true regret. The only way that I can see for such a skeptic to avoid this conclusion in the ordinary deterministic case would be to admit that there is a substantial moral difference

between manipulation cases and ordinary causal determinism—but this admission would undermine one of the major incompatibilist strategies for defending their position.

The same goes for positive corollaries of reactive moral attitudes like regret and sorrow, for instance attitudes like appreciation and gratitude. The reason why positive moral attitudes like gratitude are threatened by the demise of moral responsibility is very similar to the reason why regret and sorrow are threatened. The reason is that a central component of such attitudes is the belief in the sort of freedom required for moral responsibility—the belief that the person to whom you grateful is an apt target for praise and blame for his actions. As Galen Strawson writes, “It seems that we very much want people to be proper objects of gratitude, for example. And they cannot be proper objects of gratitude unless they can be truly responsible for what they do” (Strawson 1986, 308). Lucy Allais expresses the point similarly, saying “feeling gratitude towards someone with respect to an action involves seeing the action as flowing from her free choice” (Allais 2008, 179).

Even Pereboom concedes this point to an extent, saying, “Gratitude might well require the supposition that the person to whom one is grateful is morally responsible for an other-regarding act, and therefore hard incompatibilism might well undermine gratitude” (Pereboom 2001, 201). Pereboom says that we can still have a sense of “thankfulness” in the absence of true gratitude (a kind of thankfulness that Waller seems to equate with true gratitude), suggesting “one can also be thankful to a pet or a small child for some favor, even if one does not believe that he is morally responsible. Perhaps one can even be thankful for the sun or the rain even if one does not believe that these elements are backed by morally responsible agency” (Pereboom 2001, 201).

In my view, examples like these highlight just how far removed the attitude of “thankfulness” that we might have towards those we regard as lacking moral responsibility is from genuine gratitude. Certainly we can, as Pereboom suggests, experience joy and thankfulness when someone (or something) who lacks moral responsibility does something nice for us. But I think we want something much deeper than this out of our relationships. If the gratitude and appreciation that we can have for our dearest loved ones is diminished to the level of the kinds of emotional reactions that I can have to pets or even blind forces of nature, then it seems that something very substantial about our personal relationships has been lost.

I would also like to say a little bit about the connection between love and freedom and moral responsibility. The idea that genuine freedom and moral responsibility might be essential for love has been expressed, to different degrees, by a number of philosophers. P.F. Strawson himself claims that the range of emotions we can experience without the

moral reactive attitudes “cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other” (Strawson 1974, 10).

Some ways of arguing for this claim are misguided. For example, some, like Robert Kane, argue that the most valuable sort of love is that love which is freely chosen. And W.S. Anglin argues that love that is necessitated (whether by manipulation or coercion or the causal structure of the world) is not authentic (Anglin 1999). This idea may have some initial intuitive appeal, but it immediately runs into obvious objections. Pereboom mentions the example of familial love, such as the love between a parent and child (Pereboom 2001). It seems completely implausible to suggest that, for example, there is any exercise of will (free or otherwise) involved in the instantaneous bond of love that forms between a mother and a newborn child. In fact, it would seem inappropriate for such a bond to have to be mediated by any effort of will on the part of the mother, free or otherwise. If the mother had to actively will herself to love her new child, we would take it as a sign that something was awry. In this instance, completely unwilling, unfree love seems to be the ideal. The same can be argued for romantic love. As Nomy Arpaly reminds us, there is a sense in which we find it romantic to say “it had to be you” — to express the fact that there is no possible way I could fail to love you (Arpaly 2006).

So I think the claim that moral responsibility is necessary for love because love must be freely chosen is mistaken. Still, moral responsibility does, in my view, play an important role in grounding our loving relationships. Consider, for example, the essential role played by the emotions discussed above—gratitude, regret, sorrow, and related attitudes like forgiveness—in loving relationships. Insofar as these are an essential component of fully deep, authentic loving relationships between adults, loving relationships are deeply diminished in the absence of moral responsibility. To see this, just imagine a relationship with a person who is regularly manipulated (pick your favorite evil neuroscientist manipulator story) in ways that rob her of responsibility for her actions. She does kind things for you sometimes, other times she is thoughtless or hurtful, but in all of her interactions she is thoroughly manipulated in ways that rob her of responsibility, that make it impossible to feel deep gratitude towards her for her kindness, or for her to truly regret her bad behavior or take responsibility for it, etc. You might feel some strong affection for her, even a kind of love, but it seems to be it would be substantially diminished in comparison to that which we feel for those who we believe to be the apt targets of the reactive attitudes that comprise moral responsibility.

In this section I hope to have fairly characterized the kind of case that I think can be made in defense of the importance of moral responsibility drawing on the kinds of

considerations that Strawson first drew our attention to. Ultimately I think it is a strong case. In the next section, I want to a bit more about how far I think this case can be extended, and offer some suggestions about where its limits may lie.

V. The Limits of the Reactive Attitudes

Now that I have said a bit about the ways in which I think some pragmatic considerations—in particular connected to the nature and quality of our relationships with others—can justify and ground moral responsibility, I want to explore in a bit more detail some of the limits of this justification.

First, I want to suggest that the strength of this kind of justification varies according to the context in which someone is being held responsible for his or her actions. I think this kind of justification is strongest in the context of every day life, in our ordinary kinds of interactions with people. As I argued in the previous section, the best defenses of the moral responsibility system are connected to the role it plays in our lives and our relationships. It is essential for attitudes like gratitude and regret, sorrow and pride, attitudes that are essential for our loving relationships, and also to our regard for ourselves. In that context, it makes sense to say that people deserve the ordinary kinds of reactive attitudes and treatments (positive or negative) that their (positive or negative) treatment of others invites.

And I think that in this context, the skeptical worries that I have raised so far are at their weakest. Consider for example the standard incompatibilist arguments I sketched earlier, the ones that suggest source or leeway requirements for moral responsibility. As I argued in the last section, it seems implausible to say that love hinges on any claims about people being the ultimate source of their love, or of having any choice in the matter at all. We care that people we love be autonomous in some sense—it would seem difficult to feel genuine love, or to have a full range of moral reactive attitudes, for a thoroughly manipulated agent—but it strains credulity to suggest that love requires contra-causal freedom of the sort that incompatibilists insist on.

Likewise, I don't think that the skeptical worries raised by the difficulty of self-orchestrated moral change and moral growth pose a very strong a threat to moral responsibility in ordinary detail contexts and in our personal relationships. On the contrary, these considerations may even help in some ways to support the importance of holding people (both ourselves and others) morally responsible for their actions. One of the points I emphasized in that section is that moral change and moral growth often requires substantial input and help from others. Many defenders of moral responsibility

(especially compatibilists) have appealed to the communicative role of our moral responsibility practices. One of the ways we communicate our moral expectations to others is in our emotional expressions—in our approval or disgust or shame or gratitude or anger, etc. And through these communications, via the moral reactive attitudes that comprise our moral responsibility practices, we can help one another to grow morally (for a detailed discussion of evidence supporting this, see Shaun Nichols 2007).

In short, given that the reactive attitudes are constitutive of so much that is essential to our relationships and our self-regard, and given they have an important role to play in how we grow and develop as moral agents, skeptical arguments carry less force in this context. The Strawsonian picture is most compelling here.

I think things are a bit different, however, when we shift to a context like criminal justice. Here our judgments of responsibility and praise and blame have much more serious consequences. When talking about criminal justice and criminal punishment, the stakes are very high. When we incarcerate criminals, we deprive them of liberty and subject them to conditions that are severe impediments to living a life of any kind of quality. For severe crimes we sometimes even deprive criminals of their lives. And the justification for this sort of practice is closely tied with moral responsibility. As Stephen Morse puts it, “both the criminal and the medical-psychological systems of behavior control require a justification in addition to public safety—desert for wrongdoing or non responsibility (based on disease)—to justify the extraordinary liberty infringements that these systems impose” (Morse 2013, 29).

There are two important points to be made here. The first is that when the stakes are this high, the epistemic standards should be raised. If the justification for a criminal justice system that deprives people of liberty is going to be grounded in moral responsibility and desert, then the justification for believing that criminals in a particular instance are responsible in the sense that could ground desert must be very strong. Pereboom expresses this point as follows: “As I argued in the context of criminal punishment, if one aims to harm another, then one’s justification must meet a high epistemic standard. If it is significantly probable that one’s justification for the harmful behavior is unsound, then it is best that one refrain from engaging in it” (Pereboom 2014, 318). What Pereboom expresses here seems right. Even if the kinds of skeptical arguments I’ve discussed in this paper fall short of being decisive in the context of a criminal justice system, insofar as they raise significant doubts and lower our level of credence in our convictions about the moral responsibility of criminals, they do provide a strong reason to exercise restraint in criminal punishment. Even if we think the odds of moral responsibility skepticism being the correct view is fairly small, it still may be

reasonable to judge that the risk that we may cause great harm to people who do not deserve it (on the small chance that moral responsibility skepticism is correct) is morally significant enough to revise our criminal justice system and treat criminals somewhat more as we should if skepticism were true.

The second point is that the skeptical arguments carry much more force when considered in the context of criminal justice. As I argued before, the claim that ultimate sourcehood or leeway conditions, in the incompatibilist sense, are necessary to ground authentic love or the moral reactive attitudes in our personal relationships is not very compelling. But I think that these arguments are much more compelling when we are talking about the kind of responsibility involved in justifying deprivation of life or liberty or other seriously harmful punishments. An incompatibilist requirement like the requirement that one be the ultimate source of his or her character makes much more sense when trying to argue that one deserves something as severe as capital punishment (for example) because of their wrongdoing. Similarly, the worry that we might have a very limited capacity to shape our own moral characters without input from others, a worry that raises substantial problems of moral luck, is most pressing when we are talking about inflicting serious harm on people for the crimes that their characters drive them to commit. There is a reason that many skeptics (like Waller and Pereboom) focus heavily on questions of criminal justice and social justice when advancing incompatibilist arguments like these—because it is in these contexts that the arguments carry the greatest intuitive force.

VI. What Sort of Criminal Justice System Should We Have?

The question that remains now is what should our criminal justice system be like? What are the costs of altering or giving up (at least some) of our traditional ideas of moral responsibility and blameworthiness in the context of criminal justice? I have argued that abandoning the idea of moral responsibility in our daily lives diminishes our relationships with others and our self-esteem. But would anything comparable happen if we were to modify our criminal justice system, focusing less on the suffering that criminals might or might not deserve, and instead—as a skeptic would prescribe—more on forward looking considerations (see Pereboom 2014) like rehabilitation and crime prevention? In my view there is no strong reason to think this.

On the contrary, there are several good reasons reason to worry about a justice system that places too much emphasis on retributivism. For starters, there is the worry that a justice system that places too much emphasis on retributivism will be limited in

the extent to which it engages in investigating and learning about the causes of people's actions. This may not be a limitation that exists as a matter of logical necessity, but nonetheless, it does seem to be a common feature of highly retributivist societies with justice systems that put the main focus on making sure that criminals 'get what they deserve.' The basic worry is that the more we as a society are inclined to judge, the less we are inclined to try to understand. But when it comes to setting social policies, it is understanding—of psychology, sociology, economics, the effects of punitive and rehabilitative and other social policies—that we need. Waller offers a striking example of this extreme kind of retributivist attitude: "As the British Prime Minister, John Major called for harsher criminal justice measures, especially against juveniles: 'Society needs to condemn a little more and understand a little less'" (Waller 2011, 283).

There is also considerable evidence that justice systems that focus more heavily on retributivism—on harsh punishments, on making sure that criminals 'get what they deserve'—produce worse outcomes. Optimistic free will skeptics have highlighted much of this data. For example, the American justice system is well known to be one of the harshest in the world, and it has been argued that this is closely connected with our sense of 'rugged individualism' and belief in absolute individual responsibility for our actions (for example see Waller 2011, 282–287). Since 2002 the U.S. has incarcerated a greater percentage of its population than any other nation in the world—about 500 prisoners per 10,000 people, or 1.6 million prisoners total, in 2010 (see Guerino, Harrison, and Sabol 2012). The numbers get even higher if we include jails as well as prisons. We are one of few nations to retain the death penalty, we use 'life imprisonment' for a wide range of crimes in comparison to most other nations, and have continually expanded minimum sentencing laws and the use of 'three strike' laws. And yet there is little evidence that our continually increasing 'toughness' on crime has produced a significant deterrent effect. A major review of studies of the deterrence effect of harsh sentences found "...the studies reviewed do not provide a basis provide a basis for inferring that increasing the severity of sentences generally is capable of enhancing deterrent effects" (von Hirsch, Bottoms, Burney, Wikstrom 1999). These facts are well known, and yet there is little political will to soften or revise our sentencing guidelines—arguably because we are so driven by a need for retributive justice.

The question that remains now is what sort of criminal justice system should we have? In light of worries like those raised above, in addition to the skeptical arguments we have considered, is there *any* role for retributive considerations? I want to suggest that perhaps there still is. First, I would like to acknowledge Morse's point that if we are ever going to deprive people of liberty, we must have good moral justification. Skeptical

arguments are significant enough that we should be less punitive than we often are, less driven by the desire for revenge. We should err on the side of compassion and mercy when we are able, and focus more heavily on outcomes rather than deserts. But nonetheless, we can at the same time consistently say that criminals *do* deserve *some* level of approbation and punishment, and this sense of desert can be motivated by the practical considerations outlined above. And a grounding in some notion basic desert is important if we are to avoid the moral problems that arise from a criminal justice system grounded purely in consequentialist considerations.⁷

And as a further suggestion, I would just like to briefly mention one natural way to incorporate the kind of Strawsonian view I have defended in the context of personal relationships into a criminal justice system. This way can be found in the idea of restorative justice. Restorative justice is an approach to justice that focuses on the circumstances and needs of the victim, and the victim's relationship to the transgressor. Offenders are encouraged to take responsibility for their actions, and enter into a dialog with the victim to apologize and (depending on the nature and circumstances of the crime) to offer some way of making amends. Victims play an active role in determining what punishment the offender will receive. This approach to criminal justice resembles the Strawsonian view of responsibility, as grounded in our relationships with others, which I have defended in this paper. It avoids the abstract and extreme notion of desert that infests our criminal justice system as it exists, a notion which can lead to extreme sentencing, and which (as I have argued) is more vulnerable to skeptical worries. It recognizes that what one deserves for committing a crime in part consists in the effects on the victim and the needs of the victim, and can be shaped by one's relationship to the victim—even if that relationship is formed after the fact in the restorative justice process. Of course, much more needs to be said to defend and refine this approach to justice, and that would take me beyond the scope of this paper (see Sommers 2013 for an excellent exploration and defense of this kind of approach). But I do think that this approach at least holds promise—and it would be a way to develop our justice system in accordance with the kind of moral responsibility that I have argued can be well justified by pragmatic considerations.

7. I don't have time to explore these problems in detail here, but to give just one example, consider Saul Smilansky's argument that without *any* kind of moral desert, we would be morally required to make the lives of criminals as comfortable and enjoyable as possible—to give them 'funishment' instead of punishment (Smilansky 2011).

VII. Conclusion

In this paper I have presented arguments that suggest that there are limits to the kind of moral responsibility that can be justified by Strawsonian-style pragmatic compatibilist considerations. Some of these arguments have been admittedly somewhat briefly sketched. Still, I hope to have made a plausible case that while pragmatic considerations can arguably provide strong grounds for moral responsibility in the context of our daily lives (strong enough to resist the major skeptical worries), this strategy is much weaker when used to try to ground a harshly retributive criminal justice system—in particular one that resembles what exists in America today (as well as several other nations). In sum, the pragmatic justification I have been considering in this paper, the sort of justification that seems to provide strong grounds for regarding both others and ourselves as apt targets for moral reactive attitudes in the personal domain, doesn't seem adequate to justify the abstract and extreme concept of desert that seems to operate in the domain of justice. If we want to find a role for retribution and responsibility in justice, then I suggest that we need to reform our justice systems to more closely model the features of our personal relationships that provide a solid footing for the reactive attitudes in the first place.

References

- Allais, Lucy. 2008. "Dissolving Reactive Attitudes: Forgiving and Understanding." *South African Journal of Philosophy* 27 (3): 179–201.
- Arpaly, Nomy. 2006. *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton, NJ: Princeton University Press.
- Double, Richard. 2002. "The Moral Hardness of Libertarians." *Philo: A Journal of Philosophy* 5 (2): 226–234.
- Gill, Kathleen. 2000. "The Moral Functions of an Apology." *The Philosophical Forum* 31 (1): 11–27.
- Guerino, Paul, Paige M. Harrison, and William J. Sabol. 2012. "Prisoners in 2010." *Bureau of Justice Statistics*.
- Haji, Ishtiyaque. 1999. "Moral Anchors and Control." *Canadian Journal of Philosophy* 29 (2): 175–203.
- Hoffman, Martin. 2001. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
- Kane, Robert. 1996. *The Significance of Free Will*. New York: Oxford University Press.
- Levy, Neil. 2002. "Are We Responsible For Our Characters?" *Ethic@* 1 (2): 115–132.
- Margolis, Joseph. 2000. "Excerpts from Ishtiyaque Haji's Discussion with Members of the Audience." *The Journal of Ethics* 4 (4): 368–381.
- McKenna, Michael. 2012. "Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism." *J Ethics* 16 (2): 145–174.
- Mele, Al. 2005. "A Critique of Pereboom's 'four-case Argument' for Incompatibilism." *Analysis* 65 (285): 75–80.
- Morse, Stephen. 2013. "Common Criminal Law Compatibilism." In *Neuroscience and Legal Responsibility*, edited by Nicole Vincent, 27–52. New York: Oxford University Press.
- Mischel, Walter, Ebbe B. Ebbesen, and Antonette Raskoff Zeiss. 1972. "Cognitive and attentional mechanisms in delay of gratification." *Journal of Personality and Social Psychology* 21 (2): 204–218.
- Nichols, Shaun. 2007. "After incompatibilism: A naturalistic defense of the reactive attitudes." *Philosophical Perspectives* 21 (1): 405–428.
- Pereboom, Derk. 2001. *Living without Free Will*. Cambridge: Cambridge University Press.

- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Slote, Michael. 2010. *Moral Sentimentalism*. New York: Oxford University Press.
- Sommers, Tamler. 2013. "Partial Desert." In *Oxford Studies in Agency and Responsibility, Volume 1*, edited by David Shoemaker, 246–262. Oxford: Oxford University Press.
- Strawson, Galen. 1986. *Freedom and Belief*. Oxford: Clarendon Press.
- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75 (1–2): 5–24.
- Strawson, P. F. 1974. *Freedom and Resentment, and Other Essays*. London: Methuen.
- Van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- Von Hirsch, Andrew, Anthony Bottoms, Elizabeth Burney, and P-O Wikstrom. 1999. *Criminal Deterrence and Sentence Severity: An Analysis of Recent Research*. Oxford: Hart Publishing.
- Waller, Bruce N. 2011. *Against Moral Responsibility*. Cambridge: MIT Press.
- Watson, Gary. 2004. *Agency and Answerability: Selected Essays*. Oxford: Clarendon Press.
- Zelazo, Philip David, and Stephanie M. Carlson. 2012. "Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity." *Child Development Perspectives* 6 (4): 354–360.