Moral Responsibility and the Strike Back Emotion: 
Comments on Bruce Waller’s *The Stubborn System of Moral Responsibility* 

Forthcoming in *Syndicate Philosophy* 1 (1): July/August 2016 

Gregg D. Caruso 

In *The Stubborn System of Moral Responsibility* (2015), Bruce Waller sets out to explain why the belief in individual moral responsibility is so strong. He begins by pointing out that there is a strange disconnect between the strength of philosophical arguments in support of moral responsibility and the strength of philosophical belief in moral responsibility. While the many arguments in favor of moral responsibility are inventive, subtle, and fascinating, Waller points out that even the most ardent supporters of moral responsibility acknowledge that the arguments in its favor are far from conclusive; and some of the least confident concerning the arguments for moral responsibility—such as Van Inwagen—are most confident of the truth of moral responsibility. Thus, argues Waller, whatever the verdict on the strength of philosophical arguments for moral responsibility, it is clear that belief in moral responsibility—whether among philosophers or the folk—is based on something other than philosophical reasons. 

He goes on to argue that there are several sources for the strong belief in moral responsibility, but the following four are particularly influential: First, moral responsibility is based in a powerful “strike back” emotion that we share with other animals. Second, there is a deep-rooted “belief in a just world”—a belief that, according to Waller, most philosophers reject when they consciously consider it, but which has a deep nonconscious influence on what we regard as just treatment and which provides subtle (but mistaken) support for belief in moral responsibility. Third, there is a pervasive moral responsibility system—extending over criminal justice as well as “common sense”—that makes the truth of moral responsibility seem obvious, and makes challenges to moral responsibility seem incoherent. Finally, there is the enormous confidence we have in the power of reason, which mistakenly leads us to believe that our conscious, rational, and critically reflective selves are constantly guiding our behavior in accordance with our deep values. 

In these comments, I would like to discuss the many points of agreement I have with Waller, providing along the way additional fuel for his skeptical fire (i.e., his moral responsibility skepticism and his skeptical analysis of the source of our strong belief in moral responsibility). I will also discuss, however, my one main point of disagreement—i.e., his desire to preserve the conception of free will. Waller believes free will can “flourish” in the absence of moral responsibility (see Ch.8), while I maintain they that the variety of free will that is of central philosophical and practical importance is the sort required for moral responsibility in a particular but pervasive sense. This sense of moral responsibility is set apart by the notion of *basic desert* and is purely backward-looking and non-consequentialist (see Pereboom 2001, 2014; Caruso and Morris, forthcoming). Understood this way, the sort of free will at issue in the historical debate is a kind of power or ability an agent must possess in order to justify certain kinds of desert-based judgments, attitudes, or treatments in response to decisions or actions that the agent performed or failed to perform.
To begin, let me first acknowledge my agreement with Waller concerning the philosophical arguments for moral responsibility, which tend to be weaker than the corresponding belief philosophers have in moral responsibility. Consider, for example, Peter van Inwagen’s dogged, resolute, and (one may say) stubborn belief in moral responsibility. After championed the consequence argument in favor of incompatibilism, van Inwagen proceeds to argue that we must reject determinism even though it means free will “remains a mystery” (1983, 2000), since to “deny the free-will thesis is to deny the existence of moral responsibility, which would be absurd” (1983: 223). He then proceeds to argue that, if science were one day able to present us with compelling reasons for believing in determinism, “[t]hen, and only then, I think, should we become compatibilists” (1983: 223)—despite, of course, all his efforts defending the consequence argument.

Additional evidence of the kind of stubbornness Waller has in mind can be found among agent-causal libertarians—such as C. A. Campbell (1957), Richard Taylor (1963), and Roderick Chisholm (1982)—who are willing to embrace mysterious and “god-like” powers and abilities to preserve moral responsibility. Chisholm, for example, famously argued: “If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved” (1982: 32). As Waller so eloquently and correctly points out: “When contemporary philosophers are willing to posit miracles in order to save moral responsibility, the philosophical belief in moral responsibility obviously runs deep and strong” (2015: 3).

Compatibilists, of course, reject miracles and propose accounts of moral responsibility consistent with our naturalistic (and even deterministic) worldview, yet they seldom provide justification for the moral responsibility system itself. In lieu of justifying the moral responsibility system, compatibilists typically take the system as given and instead focus on what attitudes, judgments, and treatments are justified from within the system. P. F. Strawson (1962) is a good example of this. His defense of the reactive attitudes takes our normal moral responsibility practices as given and proceeds from there to articulate special circumstances when it is acceptable not to hold someone morally responsible or to excuse them—e.g., when they are profoundly impaired by delusion or lack any moral capacity, either temporarily or permanently. In such circumstances, we adopt what Strawson calls the objective attitude. But according to Strawson and his followers, the denial of all moral responsibility is unacceptable, self-defeating, and/or impossible, since to permanently excuse everyone would entail that “nobody knows what he’s doing or that everybody’s behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense…” (Strawson 1982: 74). The problem with this defense of moral responsibility, however, is that it takes for granted the very thing in need of justification. As Waller so eloquently explains:

[I]f we start from the assumption of the moral responsibility system (assumptions that are so common and deep that they are difficult to escape), then the denial of moral responsibility is absurd and self-defeating. But the universal denial of moral responsibility does not start from the assumption that under normal circumstances we are morally responsible, and it does not proceed from that starting point to enlarge and extend the range of excuses to cover everyone (so that everyone is profoundly flawed).
That is indeed a path to absurdity. Rather, those who reject moral responsibility reject the basic system which starts from the assumption that all minimally competent persons (all who reach the plateau level) are morally responsible. For those who deny moral responsibility, it is never fair to treat anyone as morally responsible, no matter how reasonable, competent, self-efficacious, strong-willed, and clear-sighted that person may be. (2015: 103)

This, of course, is because the basic challenge to the moral responsibility system presented by skeptics—e.g., Waller (2011, 2015), Pereboom (2001, 2014), Levy (2011), G. Strawson (1986), and myself (Caruso 2012)—does not accept the rules of that system.

Since I agree with Waller that belief in moral responsibility is stronger than the philosophical arguments presented in their favor—either because those arguments are scientifically implausible (as in the case of agent causation), or they beg the question (as in the case of Strawson and his followers), or they end up “changing the subjection” (see Waller’s discussion in Ch.2)—in searching for the roots of the belief in moral responsibility, we must dig deeper than philosophical arguments. I also agree with Waller that the source of the strong belief in moral responsibility stems in large part from (a) our “strike-back” emotion, (b) the deep rooted belief in a just world, (c) the pervasiveness of the moral responsibility system that makes the truth of moral responsibility seem obvious, and (d) our overconfidence in the powers of reason.

Since I have already discussed the connection between just world belief and beliefs about free will, moral responsibility, and just deserts at great length elsewhere (Caruso 2014a, 2014b, 2016a; see also Carey and Paulhus 2013; Nadelhoffer and Tochetto 2013), and since my (brief) comments on Strawson above have already highlighted the power of the moral responsibility system to obfuscate the fundamental question regarding the justification of the system itself, I will limit my focus here to Waller’s discussion of the “strike back” emotion.

It is important to acknowledge that human beings share a powerful strike back emotion with other animals. When we are wronged, and when we observe another being wronged, we feel a strong and immediate urge to strike back. According to Waller, this strike back emotion is one of the main sources of our strong belief in moral responsibility:

The deepest roots of our commitment to moral responsibility are in powerful emotions, rather than reason. There are many sources for the stubborn belief in moral responsibility, and some are quite subtle. But the most basic source has the subtlety of a barroom brawl, a back-country feud, or rats locked in a frenzied death struggle: the strike-back desire when we are harmed. (2015: 39)

He goes on to add:

The vengeance motive is powerful, revenge is sweet, and retribution feels righteous. The desire to strike back, to take arms against a sea of troubles, to take revenge: this is not only a powerful desire, but one that feels morally justified. We like to punish, and we are willing to sacrifice in order to do so (Fehr and Gachter 2002; Haidt 2012, 178-179). (2015: 39)
This emotional source of our belief in moral responsibility is strong, pervasive, and (I would argue) often counterproductive with regard to achieving certain desired ends such as future safety, reconciliation, and moral formation (see Pereboom 2014).

Neil Levy, for example, does an excellent job articulating how our moral emotions tend to fuel retributive impulses, which in turn often leads to excessively punitive forms of punishment:

Human beings are a punitive species. Perhaps because we are social animals, and require the cooperation of others to achieve our goals, we are strongly disposed to punish those who take advantage of us. Those who ‘free-ride’, taking benefits to which they are not entitled, are subject to exclusion, the imposition of fines or harsher penalties. Wrongdoing arouses strong emotions in us, whether it is done to us, or to others. Our indignation and resentment have fuelled a dizzying variety of punitive practices – ostracism, branding, beheading, quartering, fining, and very many more. The details vary from place to place and time to culture but punishment has been a human universal, because it has been in our evolutionary interests. However, those evolutionary impulses are crude guides to how we should deal with offenders in contemporary society. (2016)

Crude indeed! As Waller notes: “Looking carefully at the strike-back emotion we share with rats and chimps prompts doubts of its legitimacy as a foundation for our moral thoughts” (2015: 43). When we do look carefully, what we find is that the powerful strike-back emotion overwhelms careful reflection—the kind of careful reflection that is required if we wish to adopt more effective and humane policies regarding punishment.

This is not to say, of course, that our moral emotions are always bad or that we should wish to eliminate them completely. Waller correctly points out that in certain circumstances anger provides an important ethical need (2015: 45-51)—e.g., exhibiting the right emotion when someone I love is seriously wronged. In fact, there are many emotions we do not wish to eliminate, but that we do not always regard as reliable guides to behavior. These considerations lead Waller to conclude:

Thus as a moral responsibility abolitionist I feel anger at cruel acts, and do not think it desirable to eliminate such emotions (to put it badly, I am glad I feel such angry emotions); but that is consistent with believing that it would be wrong to use those emotions as guides to behavior or as justification for the system of moral responsibility. (2015: 49)

On these points, I agree. But I would like to recommend two helpful supplements to Waller’s account.

First, like Waller, I acknowledge that the emotional reactions associated with the desire to strike back are natural, but at the same time challenge the claim that they are justified. Consider, once again, the reactive attitudes of resentment, indignation, blame, and moral anger. Since these reactive attitudes can cause harm, they would be appropriate only if it is fair that the agent be subject to them in the sense that she deserves them. We can say, then, that an agent is
accountable for her action when she deserves, in the basic desert sense, to be praised or blamed for what she did—i.e., she deserves certain kinds of desert-based judgments, attitudes, or treatments in response to decisions or actions she performed or failed to perform, and these judgments, attitudes, or treatments are justified on purely backward-looking grounds and do not appeal to consequentialist or forward-looking considerations, such as future protection, future reconciliation, or future moral formation.

The version of free will skepticism I defend, which includes a skepticism about moral responsibility (more on this in a moment), maintains that agents are never morally responsible in the basic desert sense, and hence expression of resentment, indignation, and moral anger involves doxastic irrationality (at least to the extent it is accompanied by the belief that its target deserves to be its recipient). Of course one could ask, as surely a Strawsonian would, “But can we ever really relinquish these reactive attitudes? And would it be desirable if we could?” In response, I would first say that the moral anger associated with the reactive attitudes of resentment and indignation is often corrosive to our interpersonal relationships and to our social policies (see Caruso 2016a, 2016b). Like Pereboom (2001, 2014), I contend that the expression of these reactive attitudes are often suboptimal as modes of communication in relationships relative to alternative attitudes available to us—e.g., feeling hurt, or shocked, or disappointed.

On the question of whether it is possible to relinquish these reactive attitudes, my answer begins by first distinguishing between what Shaun Nichols calls narrow-profile emotional responses and wide-profile responses (Nichols 2007; see also Pereboom 2014). It is this distinction that I offer up as a supplement to Waller’s account. Narrow-profile emotional responses are local or immediate emotional reactions to a situation. Wide-profile responses are not immediate and can involve rational reflection. I believe it is perfectly consistent for a free will skeptic to maintain that expressions of resentment and indignation are irrational and still acknowledge that there may be certain types and degrees of resentment and indignation that are beyond our power to affect. If, for example, some serious moral wrong were done to my wife and daughter, I doubt I would be able to keep myself from some degree of narrow-profile, immediate resentment (nor, as Waller points out, would I be judged kindly if I did). Nevertheless, in wide-profile cases, we do have the ability to diminish or even eliminate resentment and indignation, or at least disavow it in the sense of rejecting any force it might be thought to have in justifying harmful reactions and policies (see Pereboom 2014). And since the wide-profile emotional reactions are most important when it comes to public policy—waging war, criminal sentencing, justifying punishment, etc.—I do believe philosophical arguments against moral responsibility can change our practices and reactions.

My second supplement to Waller’s account draws on recent empirical work in social psychology, which indicates that how we assign responsibility is correlated with prior judgments of what counts as being morally bad, which are in turn dependent upon other, larger, social and cultural factors (see Hardcastle, forthcoming). Take, for example, Mark Alicke’s culpable control model of blame. It proposes that our desire to blame someone intrudes on our assessments of that person’s ability to control his or her thoughts or behavior (Alicke 2000, 2008, Alicke et al. 2008a, 2008b). As Valerie Hardcastle describes:
Deciding that someone is responsible for an act, which is taken to be the conclusion of a judgment, is actually part of our psychological process of assessing blame. If we start with a spontaneous negative reaction, then that can lead to our hypothesizing that the source of the action is blameworthy as well as to an active desire to blame that source. This desire, in turn, skews our interpretations of the available evidence such that it supports our blame hypothesis. We highlight evidence that indicates negligence, recklessness, impure motives, or a faulty character, and we ignore evidence that suggests otherwise. In other words, instead of dispassionately judging whether someone is responsible, we validate our spontaneous reaction of blameworthiness. (forthcoming)

In fact, as Hardcastle cites, data suggests that we often exaggerate a person’s actual or potential control over an event to justify our blame judgment and we will even change the threshold of how much control is required for a blame judgment (Alicke et al. 2008; see also Alicke 1994; Clark et al. 2014; Everett et al., forthcoming; Berg and Vidmar 1975; Eften 1974; Lagnado and Channon 2008; Lerner and Miller 1978; Lerner et al. 1976; Neimeth and Sosis 1973; Schlenker 1980; Snyder et al. 1983; Sosis 1974).

A recent set of studies by Cory Clark and his colleagues (2014), for example, found that a key factor promoting belief in free will is a fundamental desire to blame and hold others morally responsible for their wrongful behaviors. Across five studies they found evidence that greater belief in free will is due to heightened punitive motivations. In one study, for instance, an ostensibly real classroom cheating incident led to increased free will beliefs, presumably due to heightened punitive motivations. In a second study, they found that the prevalence of immoral behavior, as measured by crime and homicide rates, predicted free will belief on a country level. These findings suggest that our desire to blame and hold others morally responsible comes first and drives our belief in free will, rather than the other way around.

Other researchers have found that our judgment on whether an action was done on purpose or not is influenced by our moral evaluation of the outcome of certain actions—i.e., whether we morally like or dislike it (Nadelhoffer 2006). Additional findings have found an asymmetric understanding of the moral nature of our own actions and those of others, such that we judge our own actions and motivations as more moral than those of the average person (Epley 2000). As Maureen Sie describes:

> In cases of other people acting in morally wrong ways we tend to explain those wrongdoings in terms of the agent’s lack of virtue or morally bad character traits. We focus on those elements that allow us to blame agents for their moral wrongdoings. On the other hand, in cases where we ourselves act in morally reprehensible ways we tend to focus on exceptional elements of our situation, emphasizing the lack of room to do otherwise. (2013: 283)

These empirical findings help support Waller’s argument concerning the role the strike back emotion plays in our moral responsibility beliefs and practices. It appears that our moral responsibility practices are often driven, possibly primarily driven, by our desire to blame, punish, and strike back at moral transgressors, rather than, and often in lieu of, our more rational and objective judgments about free will, control, and moral responsibility.
Keeping in mind, then, that I share with Waller both his long-standing skepticism about moral responsibility and his analysis of why the belief in moral responsibility is so stubborn, I will now turn to our one point of substantive disagreement: whether or not the concept of free will should be preserved. While I completely agree with Waller that backward-looking moral responsibility, praise and blame, and the reactive attitudes of resentment, indignation, guilt, and righteous anger cannot be justified in a naturalist world devoid of miracles, I see no justification for, or benefit in, preserving his restorative notion of free will. In both The Stubborn System of Moral Responsibility (Ch.6) and in his new book Restorative Free Will (2016), Waller argues that free will can flourish in the absence of moral responsibility. Since I have recently criticized this aspect of Waller’s account elsewhere (Caruso 2016c), I will keep my comments here brief.

On Waller’s account, free will amounts to the ability to discriminate among and evaluate alternatives and the ability to adjust the level of behavioral variability to environmental conditions. I contend that this conception of free will makes the dispute between compatibilists and incompatibilists a moot point since no one in the debate denies that we have the kinds of abilities discussed by Waller. The question most philosophers are interested in—the question that is of central philosophical and practical importance in the free will debate—is whether these abilities are enough to justify certain kinds of desert-based judgments, attitudes, or treatments in response to decisions or actions that an agent performed or failed to perform. On this point, Waller and I both agree that the answer is no. It’s hard to see how Waller’s conception of restorative free will—divorced as it is from moral responsibility—helps resolve that debate, or frankly any other significant debate related to the historical problem of free will. Even if we grant Waller his restorative free will, it is difficult to think of anything of importance that follows from it regarding our everyday practices, judgments, and attitudes. By liberating free will from moral responsibility, Waller has seemingly liberated it from all of its philosophical and practical importance.

I have elsewhere argued that there are several distinct advantages to defining free will in terms of the control in action required for basic desert moral responsibility (see Caruso and Morris, forthcoming): (a) it provides a neutral definition that virtually all parties can agree to—i.e., it doesn’t exclude from the outset various conceptions of free will that are available for compatibilists, libertarians, and free will skeptics to adopt; (b) it captures the practical importance of the debate; (c) it fits with the commonsense (i.e., folk) understanding of these concepts; and, perhaps most importantly, (d) rejecting this understanding of free will makes it difficult to understand the nature of the substantive disputes that are driving the free will debate. Waller’s conception of free will, it seems to me, fails to have any of these virtues (see Caruso 2016c). I therefore encourage my good friend to follow me down the “sinful path of free will eliminativism” (Waller 216: x).

Let me end with some final thoughts. No one has influenced my thinking on moral responsibility more than Bruce Waller. For that I owe him a great debt. Like Waller, I believe we should “destroy moral responsibility, drive a stake in its heart, and bury it at the crossroads” (2016: viii). But given how strong and stubborn the belief in moral responsibility is, this will not be easy. Furthermore, Waller’s desire to preserve free will, contrary to his good intentions, may actually be standing in the way of achieving that end. Jasmine Carey and Delroy Paulhus (2013) have
recently found that where belief in free will is strongest we tend to find increased retributive moral judgments. More specifically, they found that free will believers were more likely to call for harsher criminal punishment in a number of hypothetical scenarios. Shariff et al. (2014) have reported similar findings. In one study Shariff and his colleagues found that people with weaker free will beliefs endorsed less retributive attitudes regarding punishment of criminals, yet their consequentialist attitudes were unaffected. In a different study they found that experimentally diminishing free will belief through anti-free-will arguments diminished retributive punishment, suggesting a causal relationship. This research provides prima facie support for thinking that the folk conceive of free will as linked with moral responsibility and, more specifically, retributivist judgments. If Waller wants to reduce the latter, as I know he does, his free will preservationism may be counterproductive.

References


