CAN WE RESPONSIBLY REJECT MORAL RESPONSIBILITY?: A CRITICAL ASSESSMENT OF BRUCE WALLER’S AGAINST MORAL RESPONSIBILITY

William A. Rottschaefer
Lewis and Clark College

ABSTRACT: I critically review Bruce N. Waller’s Against Moral Responsibility (2011), in which he argues that the moral responsibility system can and ought to be abandoned and that it be replaced with a take-charge responsibility system. Taking the perspective of a friendly naturalistic critic who embraces Waller’s naturalism, I argue that Waller has failed to establish his negative argument and that he has not made clear how his positive proposal differs from a recommendation to reform the moral responsibility system. Key words: determinism, free will, compatibilist free will, libertarian free will, moral responsibility, naturalism

I suspect that few people would quarrel with the proposal that the practices of slavery and sexism ought to be eliminated. But should the practice of assigning moral responsibility (MR)? Come on! Yes, indeed. Bruce Waller in his very challenging and stimulating Against Moral Responsibility vigorously and passionately argues that what he calls the moral responsibility system (MRS) is fundamentally morally corrupt and that it must be resisted and—though it will be no piece of cake—eventually destroyed. This project, he maintains, is both desirable and socially and personally possible. Moreover, that destructive effort will foster the constructive effort of building positive moral relations that enhance individual and collective lives.

Waller’s project does not stand entirely in isolation from the ongoing efforts of philosophers to delve more deeply into the very ancient questions concerning fate, determinism, free will (FW), and MR. But his is not, as he puts it, an internal question about how to clarify, refine, and improve the MRS. Waller is not attempting to answer more carefully and correctly questions about when and how to assign MR in a more morally responsible way. Revision (e.g., Vargas, 2007) does not offer a path forward. His is an external question about whether it is ever morally justifiable to assign MR. And Waller’s response to that question is loud and clear—“Never!” Yet Waller is not among the hard determinists that find FW

AUTHOR’S NOTE: I wish to thank my colleague John Fritzman for helpful comments and discussion on an earlier draft of this review and the editor of this journal, Jay Moore, for inviting me to write it. Please address correspondence to William A. Rottschaefer, Department of Philosophy, Lewis and Clark College, Portland, OR 97219; Email: rotts@lclark.edu
incompatible with determinism, even a determinism that is scientifically construed. These hard determinists reject MR and make their case against it through arguments for the rejection of FW: no FW, no MR. Waller, however, supports a scientifically-based account of FW, thus placing himself in the compatibilist camp on the FW question. But, unlike compatibilists, Waller will have no truck with either libertarian or compatibilist support of MR. Thus, Waller finds himself in a dialectically distinct position: one that supports a compatibilist account of FW but rejects MR.

His rejection contrasts with another recent eliminativist stance on MR urged by Tamler Sommers (2012). Sommers retains the traditional link between FW and MR. But Sommers argues that substantive evidential weight should be given to the phenomenon of differing individual and cultural intuitions about FW and MR. So he argues that there is, and can be, a plurality of positions on the questions of both FW and MR. In the end, after a good deal of soul searching, Sommers finds that his intuitions lead him to argue for eliminativism. Waller, on the other hand, pushes a stronger scientific naturalist line, arguing that the MRS cannot be supported scientifically and is fundamentally unfair. Thus, Waller aims to move beyond the differing intuitions that engender differing requirements for MR. He contends that scientific theories and findings enable us to hold plausible views about the existence of moral values and FW, but they make any case for MR implausible.

Nor does Waller concede that even though the MRS is fundamentally morally unjustified, it brings pragmatically beneficial consequences that might justify its retention as, for instance, a beneficial illusion (Smilansky, 2000). There is nothing that supports its retention and everything to support its abrogation. Indeed, attempts to refine the MRS often obfuscate and turn off further investigation, while the scientific findings that support the eliminative task bring with them the tools and know-how for constructing an alternative that is both personally and socially morally enhancing.

Given this rough account of the current lay of the land and Waller’s location in it, I shall now hone in on the details of Waller’s position. First I lay out the main lines of his negative argument against MR. Taking the stance of a friendly scientific naturalistic critic, I evaluate that argument in the next section of this paper. Though I find myself in agreement with its scientifically informed naturalistic approach to which he appeals, I argue that Waller’s negative argument for the rejection of the MRS fails. In the following section I discuss and evaluate his positive argument for a scientifically informed successor: take-charge

1 Waller does not explicitly address the role of intuitions in philosophical methodology. He stresses the naturalist credentials of his view and makes use of many scientific findings in his arguments. However, his book is full of often strikingly formulated stories—some of the fanciful sort that philosophers are want to construct—that some readers might be inclined to read as intuition pumps, especially when they are concluded with rhetorical questions. Yet a more charitable reading might be that they are intended only as attempts to make vivid established empirical findings. I shall return to this issue of the nature of Waller’s naturalism and the kinds of support he gives for his eliminativist account.
responsibility (TCR). I maintain that though a scientific naturalist should find his argument for TCR plausible, these findings also support a MR system and that Waller has not shown how the two systems are fundamentally distinct. Thus, his argument for the former and the rejection of the latter appears to be, despite his multiple claims to the contrary, a plea for reform of the MRS. In the end, Waller may find my critique just another manifestation of the almost unbreakable grip of the “strike-back reaction” that he discerns at the core of the MRS, a reaction to which we humans seem all too prone to rely on in our dealings with others. If Waller proves correct, I hope that my assessment at least manifests a few significant links in the chain that, in Waller’s view, must be smashed in the end.

The Argument for Eliminating Moral Responsibility

Waller presents the main lines of his argument for eliminating the MRS and for adopting a take-charge account of responsibility in the Preface and Chapters 1–3 of his book. But, in a real sense, the story has only begun. For Waller takes the reader through many of the recent arguments—both libertarian and compatibilist—for MR, showing how, in his view, they fall prey to the charge of fundamental unfairness. At the same time he fills in the picture of his alternative of compatibilist FW, moral agency, take-charge MR, and morally appropriate responses to those acting well or badly. Here I can only briefly summarize.

In Chapter 4 Waller rejects authenticity accounts of MR that link it to FW conceived not as the ability to do otherwise but as the capacity to embrace the choices one has made. However, he finds authenticity accounts of FW naturalistically appealing. Chapter 5 is devoted to rejecting both libertarian and compatibilist views of FW that find room for FW and an associated MR in the gaps left open by still scientifically unexplained aspects of free actions. These accounts suffer from the same sort of fate as attempts to find proofs for the existence of God in phenomena yet to find scientific explanation. Waller examines in Chapter 6 the efforts to account for MR as something that is taken on rather than

---

2 Upon my initial reading of Waller’s book I had the impression that there was much unnecessary repetition of material. Since philosophers are trained to be clear and concise and to avoid repetition, I found this reading tough going. However, as I have reread his book and studied it more carefully, I came to believe that my initial take on its presentation of material was not the proper one. Waller does introduce the reader to a cast of fictional characters that reappear throughout the book. But the scenes in which they act are mostly different ones. And it is true that Waller does not present the views of the major philosophical players in a single place nor once and for all in a detailed fashion. This can give the reader the impression that Waller is repeating positions that have come up before and that should have been laid out once and for all in a systematic fashion. This impression may have some validity, but I think that the better way to understand what Waller is doing is to view exponents of various positions entering and leaving the stage and then returning as the dramatic flow follows an underlying dialectic. My message to readers, especially philosophers, is: do not get annoyed by what appears to be needless repetition. Something of more importance is going on.
possessed. He finds the notion of TCR plausible, but that of MR to be unacceptable since it requires reactions of praise and blame that he argues are fundamentally unfair. Chapter 7 is devoted to the take-charge-of-oneself version of this approach to MR. In Chapter 8 Waller rejects pragmatic justifications of the MRS. In Chapter 9 Waller takes up a perhaps deeper challenge to his attempt to break the link between MR and FW. Earlier versions of the link clearly separate moral action and moral reaction. A moral agent acts well or badly, and she herself or another reacts to that action with either praise or blame. The praise or blame, it is argued, is justified because the agent is morally responsible, be that responsibility one that is possessed or taken on. The justification for the reaction is provided by moral normative arguments for its fairness or pragmatic arguments for its benefits. But some proponents of MR find a tighter connection between FW and MR. Taking character to be the result of free actions and knowing free choices—something that Waller roughly accepts—they argue that the attribution of character faults are necessarily attributions of MR and are manifested in various negative responses, in particular blame. Waller finds this view fallacious, maintaining that it fails to distinguish between being at fault and deserving blame for being at fault. People can genuinely be at fault, but not necessarily responsible. And blame for fault is always unfair. Chapter 10 finds Waller rejecting the argument that without MR the entire moral life disappears, arguing that morally good and bad actions and character, as well as appropriate reactions to them, remain. Only the unfair practices of praise and blame and reward and punishment are eliminated. In Chapter 11 Waller distinguishes the MRS system that engages in excuse-extension from his own naturalistic TCR system that rejects “miracles.” Chapter 12 explicitly addresses compatibilist accounts of MR that see the latter as a kind of plateau. Waller rejects these views because they make a moral agent responsible for factors over which she has no control—factors that are matters of luck. Arguments that attributions of MR are required as a matter of respect are rejected in Chapter 13. Chapter 14 finds lacking arguments that situate the need for MR within the phenomenon of genuine authorship. Chapter 15 describes a world without MR, and Chapter 16 finds Waller arguing vigorously for the possibility and desirability of bringing about such a world.

Waller’s Major Claims

Waller succinctly sets out his major claims in the Preface to his book:

(1) The MRS cannot survive a scientifically-based naturalistic critique.

(2) A naturalistic responsibility system (associated with what he calls take-charge responsibility, or TCR), distinct from the MRS, remains after the critique of the latter.

(3) It is possible, desirable, and morally required to reject the MRS and to take on and develop a naturalistic responsibility system.
The MR of the MRS is the “moral responsibility that justifies special reward and punishment” (p. 2). Waller refers to this conception of MR as “our basic notion” (p. 5) and “the core concept” (p. 7). He goes on in this chapter to mention alternative conceptualizations of MR such as (1) a social benefit view, (2) as equivalent to making a moral judgment, and (3) an accountability view which he critiques in this chapter. He concludes:

Whatever the merits or faults of other views of moral responsibility, this book focuses on what I take to be the core concept of moral responsibility: moral responsibility is what justifies blame and praise, punishment and reward; moral responsibility is the basic condition for giving and claiming both positive and negative just deserts. (p. 7)

He continues:

My goal in this book is to show that claims and ascriptions of moral responsibility (in the robust sense specified previously) cannot be justified, that there are strong arguments to show that—absent miracles—the system of moral responsibility and “just deserts” is fundamentally unfair, and that we will be better off when belief in moral responsibility is utterly eliminated. (p. 7)

We will need to return later to the issue of how Waller understands his project. What does Waller mean by a “core” concept? Does he have in mind a folk-psychological notion or a biologically-based notion? Is there such a core concept, and what would it mean to say that there is? If there is, has Waller identified the core concept of MR? And what counts as elimination? Pictures of eliminativism conjure up quintessence, the aether, phlogiston, and witches. Getting rid of these conceptions was, to put it paradoxically, getting rid of nothing. But Waller links MR to a basic irrational retributive impulse. Prescribing its elimination does not seem to fit the pattern of the sort of theoretical cognitive elimination that the classical eliminativist position seems to typify. It sounds more like a recommendation for behavioral change. And since, as we shall see, the argument against the MRS is basically a normative one, it appears to be one about fundamental moral change. The elimination of MRS does not appear to be merely about discovering that what one has been talking about, and characterizing refers to nothing and does not exist. Rather, it appears to be a discovery about oneself and one’s behavior and about a significant source of one’s behavior. Moreover, it is a realization that this behavior has causal consequences of a moral sort that ought not to be. Rather, they ought to be replaced by something morally worthy. We find out that we ought to stop doing what we are doing. We must eliminate it as we did slavery and sexism. A mere cognitive or theoretical enterprise this sort of

---

3 Unless otherwise noted, all page references are to the volume under review.
4 Sommers documents significant cross-cultural and cross-temporal diversity in concepts of MR. Waller adverts to some of the cross-temporal development of and diversity in what he calls the basic retributive emotion.
eliminativism is not; rather, it is primarily a moral project to eliminate a corrupt moral practice.

So what is this corrupt system? MRS is a complex system consisting of cultural, social, and individual practices of praising and blaming and punishing and rewarding. These practices bring with them an arsenal of attempted justifications including, most importantly, diverse philosophically-based justifications. But, on Waller’s account, those justifications are themselves founded on a basic belief in the system that has its roots in an emotional commitment, characterized by Waller as an irrational retributive impulse (pp. 7-14). Waller contends that that impulse has a dark evolutionary history, one that clearly reveals why it ought to be rejected. It is, of course, another question about whether, granting it an evolutionary status, it can be uprooted. However, it would seem that Waller’s project does not require genetic engineering for its accomplishment. Given the immorality of attributions of MR, we can work to change the social environment and to change our social cognitive and behavioral practices. These efforts can render the retributive emotion and its attendant attitudinal and cognitive structures, as well as its would-be justifications, relatively inoperative.

In the end Waller finds the philosophical attempts to justify MRS by appeals to libertarian ideas of FW or compatibilist notions that link compatibilist FW with suitably modified accounts of MR to be, at best, failed attempts to prop up a basic belief in MR that itself gets its footing from that dark impulse. Waller makes use of social psychologist Jonathan Haidt’s (2001) view that moral reasons find their basis in emotions. And, for that reason, moral justification is usually a kind of legalistic rationalization of previously held emotionally-based beliefs (p. 94). Thus, he maintains that these cognitive endeavors toward justification may be nothing more than rationalizations for independently entrenched irrational commitments.

The Unfairness Argument Against Moral Responsibility in Its Dialectical Context

Waller succinctly formulates what he calls the basic argument against MR—the comparative unfairness argument—in this way:

The fundamental naturalistic argument against moral responsibility is that it is unfair to punish one and reward another based on their different acts because their different behaviors are ultimately the result of good or bad luck. (p. 22)

We can place this argument within a straightforward dialectic of arguments for and against FW. On the pro side we have:

P1. Moral responsibility requires free will.
P2. We have moral responsibility.

5 I use the formulation of “having moral responsibility” to keep the arguments straightforwardly valid. I avoid the complications that a more precise formulation such as “we ought to be morally responsible” entails. I do not believe that my simplified versions
C1. We have free will.

In its libertarian version we have:

LP1. Moral responsibility requires libertarian free will.
LP2. We have moral responsibility.
LC1. We have libertarian free will.

In its compatibilist version we have:

CP1. Moral responsibility requires compatibilist free will.
CP2. We have moral responsibility.
CC1. We have compatibilist free will.

Both libertarians and traditional compatibilists accept P2, maintaining that we have MR. They differ on P1, concerning the kind of FW that MR requires. Libertarians maintain that FW is incompatible with determinism, so they argue that humans must possess a kind of causal capacity that enables them to choose between alternatives or not to choose at all—a capacity that makes them independent sources of action. On the other hand, traditional compatibilists hold that FW and determinism are compatible. Humans are sometimes able to act in ways that are neither physically nor psychologically compelled. Hard determinists deny both FW and MR. They accept the connection between FW and MR, thus accepting P1 in either version. But on the basis of determinism they deny the existence of either libertarian or compatibilist FW. Thus they argue:

DP1. Moral responsibility requires free will.
DP2. Neither libertarian nor compatibilist free will exist.
DP3. We do not have moral responsibility.

Consequently, both libertarians and traditional compatibilists maintain that the practices of praising and blaming, rewarding and punishing are morally proper because humans are morally responsible, though, of course, such practices must be tempered by the degree of responsibility had by the agents in question. Compatibilists do more tempering than libertarians. Determinists, on the other hand, reject the practices of praising and blaming because they contend that the existence of determinism precludes the existence of any sort of FW, whether libertarian or compatibilist.

Waller denies DP2, arguing for the existence of compatibilist FW. But he also rejects the claim of MR enunciated in P2 and instantiated in LP2 and CP2. Waller

prejudice any central issues. The simplified and more complex versions all embrace the key idea that some sort of FW is a necessary condition for moral responsibility and the latter is a feature of human behavior.
formulates his argument against MR in terms of what he calls the comparative unfairness argument, but he maintains that formulation

...is just another way of presenting the same basic argument against moral responsibility, or—more precisely—the same basic argument to show that claims and ascriptions of moral responsibility to humans (humans who lack godlike miraculous powers of originating self-creation) are unfair. (pp. 23-24)\(^6\)

**Reconstructing Waller’s Argument**

Waller does not present his argument formally. Rather, he proceeds by confronting the many and varied arguments of his opponents with a variety of empirically-based claims and vivid and engaging examples to illustrate these claims and his counter-arguments. Working from these examples and his informal presentations, I reconstruct in a more formal fashion his argument for the elimination of the MRS, adding in parentheses a summary support or characterization for each premise.

W1. Moral responsibility is a feature of a moral agent’s moral decisions, actions, and character that requires a response from the agent herself and from other moral agents. (Agreed-upon description of the phenomenon under investigation)

W2. The response is either one of praise or blame, reward or punishment. (Agreed-upon summary account of the required responses evoked by the moral responsibility feature of moral agents)

W3. The moral responsibility constituting feature requires a crucial causal control by the moral agent of her actions, decisions, character, and any other factors that make for any aspect of one’s moral self—moral self-features, for short. Call this the moral responsibility–FW connection—the MR-FW connection. It comes in two varieties: libertarian and compatibilist. (A commonly held view of libertarian and compatibilist proponents of moral responsibility)

W3L. On the libertarian account, the practices of praising and blaming and of rewarding and punishing require a moral agent’s crucial causal control to be

---

\(^6\) Though he focuses mostly on libertarian notions of FW in Chapter 2 while formulating his argument, Waller finds compatibilists’ reduction of the requirements for FW, and thus for moral responsibility, no more convincing. On Waller’s view, compatibilists, especially those who have paid attention and employed scientific findings, have forged an acceptable view of FW. Their failure is to remain attached to the illusion of moral responsibility. The latter, unlike the former, is an all-or-nothing affair. We will need later to explore in some more detail the asymmetry of Waller’s views.
*complete* causal control over her entire moral self. (The libertarian version of the MR-FW connection)

W3C. On the compatibilist account, the practices of praising and blaming and of rewarding and punishing require a moral agent’s crucial causal control to be an appropriate natural or acquired causal control over one’s moral self. (The compatibilist version of the MR-FW connection)

W4. The practices of praising and blaming and of rewarding and punishing can be *fair* only if they are responses to actions over which a moral agent has crucial causal control. (Assumed agreement concerning a *guiding moral principle* for the proper moral assessment of the moral responsibility practices)

W5. No one has the kind of crucial causal control over their moral self that is sufficient and necessary for the attribution of moral responsibility. (The abolitionist’s general denial of moral responsibility—one supported by Waller)

W5L. Naturalism establishes that the God-like powers or miracles required by the libertarian MR-FW connection do not exist. (Denial of libertarian moral responsibility)

W5C. Naturalism establishes that moral agents have only some of the necessary causal conditions for the compatibilist MR-FW connection. (Denial of compatibilist moral responsibility)

W6. Therefore, the practices of praising or blaming and rewarding or punishing are unfair. (Follows from W4 & W5)

W7. The MRS requires moral responsibility. (Agreed-upon essential requirement for a MRS)

W8. Consequently, the MRS ought to be eliminated. (Follows from W6, W7, and the obvious ethical principle that unfair practices ought to be eliminated, given that they are immoral practices)

The heart of Waller’s negative argument elaborated through the course of the book is aimed at establishing W5 in both its versions. Waller seeks to accomplish this by appealing to naturalist philosophical views as well as scientific findings. These are made vivid by illustrations of cases in which Waller asks his readers to apply their normative judgments about the fairness of attributions of MR, given either the libertarian or compatibilist version of the MR-FW connection. Waller argues, and expects the reader to determine, that neither version of the MR-FW connection provides an adequate basis for a *fair* attribution of responsibility.
Waller emphasizes the distortions in accounts of FW that the assumption of MR generates. These, of course, are most obvious in libertarian versions. Indeed, the inflated accounts of FW that these generate induce the more modest views of the compatibilists. Yet both neglect the central question generated by the attribution of MR: are the practices of praising and blaming or rewarding and punishing ever fair? The origins of these practices suggest a negative reply.

Waller’s naturalistic account of the origin of notions of MR turns to our evolutionary history. He raises the question of how such practices might have arisen and how they might have contributed to evolutionary success. He suggests that the notion of MR is rooted in an evolutionarily-based strike-back desire:

Moral responsibility is rooted much deeper than philosophy, deeper than reason. The desire—indeed, the visceral biological need—to strike back at trouble can be found in chimpanzees and rats. It existed long before humans, and long before concern with free will or moral responsibility. It is based in a profoundly irrational strike back inclination. . . .that targets wrongdoers only by chance. (p. 49)

Waller provides some scientific support for this evolutionary story, then he proceeds to sketch a brief history of the development of the notion of MR, tracing the reflective attempts to justify the practices that the strike-back desire promotes. These attempts range from appeals to divine authority to philosophical accounts that lay responsibility on the shoulders of human FW (pp. 49-53). Thus, on Waller’s account, as he suggests in Chapter 1, philosophical accounts of FW have developed as rationalizations for the evolutionarily-based, irrational strike-back desire.

Scientific naturalists should approve Waller’s move to understand the origins of the notions of FW and MR—but the adequacy of his evolutionary and cultural stories about its development is another matter. In addition, his assessment of moral justification as rationalization needs careful attention and critique. I shall return to these matters below.

**Assessing Waller’s Rejection of the MRS**

As we have seen, Waller aims his eliminativist’s sights at what he calls the core concept of MR. That concept is captured by W1, W2, and W3. It is a common assumption of both Waller and the proponents of MR whom Waller is opposing. Recall what he tells us:

> Whatever the merits or faults of other views of moral responsibility, this book focuses on what I take to be the core concept of moral responsibility: moral responsibility is what justifies blame and praise, punishment and reward; moral responsibility is the basic condition for giving and claiming both positive and negative just deserts. (p. 7)

A number of issues need sorting. First, what is a core concept? Is there a core concept of MR? If so, what is it? Does the identified core concept track any
phenomena? And do these phenomena possess normative weight? Waller does not elucidate what he means by a core concept of MR. As far as I can determine, he understands a core concept to be a commonly held understanding of a phenomenon. Though he alludes to historically and culturally differing views of MR and its bases, Waller does not seem to see these as reasons for questioning the view that there is a common understanding of MR; he assumes that there is such a common understanding of MR. On his account, philosophers attempt to capture the core concept of MR by providing correct analyses that may require some refinement, and they argue for the normative appropriateness of the MRS as well as delineating the ways in which its application needs to be suited to the details of differing moral agents and situations. On Waller’s view, that common understanding tracks the response-evocative property constituted by the activity of moral agency. Moral agents perform morally good and bad actions and become morally good or bad persons, with good or bad moral characters. These actions and characteristics provoke moral assessments. And according to the proponents of MR, these actions and characteristics morally demand responses of praise and blame, reward and punishment. But Waller does not offer any argument showing his view to be the core concept. He seems to assume that the response-evocative conception of MR, one that links it either to libertarian or compatibilist FW, along with the normative response that it entails, constitutes the core concept of MR.

What are we to make of this? First, realists (especially of the scientific naturalistic persuasion, to which Waller seems to have some affinity) might question Waller’s focus on concepts rather than on the phenomena to which these concepts are supposed to refer. Why should a proclaimed naturalist about MR be interested in concepts and their analyses? Even if Waller has correctly identified the core conception of MR, thus showing that there are, according to that notion, conceptual links between the concepts of moral action, moral judgment, and praise/blame (reward/punishment), why should this be of interest (or of central interest) to a naturalist? Waller’s method seems to be a non-naturalistic one that relies on conceptual analysis. His use of imaginary cases and appeals to intuition

---

7 Waller does not address the issue of whether such a core concept is biologically based and/or a social–cultural product. We might take his suggestion that the conception of MR is rooted in a strike-back emotion as identifying a biologically-based factor.

8 Waller contrasts his own understanding of the core concept of MR with several alternatives. These are what he calls the social benefit view, the moral judgment view, and the accountability view. He critiques the social benefit view in Chapter 8 as normatively inadequate and failing to live up to its own standards. And in Chapter 10 he maintains that his own positive account of TCR retains both moral phenomena and the moral judgments that it evokes, but he rejects as unjust the normatively required practices of praise and blame, reward and punishment that, according to MR, these moral judgments require. Thus, Waller seems to be implicitly arguing that a moral judgment account of MR does not capture the core concept because it leaves out these practices. In Chapter 1 Waller rejects the accountability view because people often give incorrect accounts of why they acted as they did and, even if they correctly identify the reasons for their actions, these reasons do not establish that they are morally responsible for their actions.
reinforces this impression. Though Waller proclaims his position to be naturalistic and though he takes a substantively naturalistic position having no truck with god-like powers and miracles, his approach has elements of what has come to be called “first philosophy,” where analyses of concepts are sufficient for making claims about phenomena. Thus, scientifically-inclined philosophical naturalists who embrace both methodological and substantive naturalism would argue that even if Waller’s critique of MRS turns out to be correct as an exercise in conceptual analysis, it might tell us nothing about the natural realities which are the supposed referents of these concepts. However, it seems clear that Waller’s intention is to deal with the actual phenomena and with empirically-based connections rather than with mere concepts and the connections revealed by conceptual analysis. And his appeal to scientific findings and theories indicates that he intends to approach the issue of MR in a methodologically naturalistic fashion. So, his claims that there is a core concept of MR and that it is as he describes it are substantial empirical claims, ones that need empirical support. And, as it turns out, there is a substantial amount of interesting and relevant recent empirical work that has been done on the issue. So it is possible to assess the empirical adequacy of Waller’s claims not only about the content of the core concept of MR but, indeed, whether there is such a core concept. In discussing these findings I follow the very helpful summaries of the recent empirical research put together by Joshua Knobe and John M. Doris (2010) and by Doris, Knobe, and Woolfolk (2007).

Many philosophers have held that the common-sense conception of MR-FW connection is incompatibilist. Assuming a MR-FW connection, most philosophers have thought that ordinary folk believe that in a deterministic world there can be no FW or MR. However, as happens frequently with the a priori assumptions of philosophers, that assumption has been shown to be problematic. Philosophical presuppositions about what ordinary people’s conceptions of the relationships are between MR and FW, on the one hand, and determinism, on the other, have inspired experimental philosophers to do some empirical testing. Empirical findings show that some folk in some situations sometimes have compatibilist views about MR and determinism. These findings have led to disputes among philosophers as to which view common folk really have. Of course, this descriptive question (what concept(s) are people using when they talk about MR?) needs to be distinguished from the epistemically normative question (what phenomenon or phenomena, if any, are their conceptions tracking?). And both questions need to be distinguished from the morally normative question (what sort of response is appropriate for the tracked reality?).

Before we take a look at some of the empirical findings, let’s get a fix on Waller’s answers to these questions, given what our examination of his views has told us. On the descriptive question, Waller assumes that there is a core concept of

---

9 This version of first philosophy reflects the resurgence of the ambitious sort of analytic philosophy of the Kripke and post-Kripkean kind, which has metaphysical ambitions, not the ordinary-language approaches of an earlier Wittgensteinian period that had disdained such ambitions.
MR, one that connects the practice of praising and blaming with what we have
called in Premise W3 “crucial causal control.” However, we have also seen that
Waller argues in answer to the epistemically normative question that there is no
response-evoking feature connecting MR to either libertarian or compatibilist FW.
Thus he maintains that the folk should be neither MR incompatibilists nor MR
compatibilists. The folk track nothing real and thus (in answer to the morally
normative question) the use of either conception in eliciting a response results in a
gravely unfair and unjust set of practices.

Knobe and Doris (2010) have examined the philosophical debates concerning
the folk conception of MR and have concluded that they have ended in stalemate.
Their diagnosis for this failure is that the philosophical method of proposing
hypothetical cases to test a priori intuitions about people’s conceptions of MR is
fatally flawed methodologically. What is needed is empirical investigation. In
addition, they maintain that recent empirical research indicates that the
presupposition of both compatibilist and libertarian proponents of MR that there is
a set of common criteria for assigning responsibility is problematic. Thus, if recent
empirical research holds up, Waller’s attack on the core concept is misdirected,
for the empirical evidence suggests that ordinary people are using a range of
conceptions in their ascriptions of MR and their reactions to it. Acceptance of
either the libertarian or compatibilist conception of the folk view and the supposed
phenomena it is tracking is problematic. Premise W3 in the master argument is
false.

The situation of the folk is much more complex than proponents of a common
set of criteria for attributing MR have supposed. In discussing these findings,
Knobe and Doris (2010) distinguish between invariance and variance theories of
MR. The former hold (as does Waller) that the answer to the descriptive question
provides a set of commonly held conceptual criteria for the identification of MR.
The latter argue that these criteria vary with a number of factors. Thus, invariance
theories postulate that people apply the same basic criteria in all cases when
making MR judgments, even though these need not be a set of necessary and
sufficient factors. Variance theories, on the other hand, maintain that the factors
inducing MR attributions vary with types of cases.

In surveying the literature, we find a number of studies that support variance
theories. These studies have a common structure that demonstrates that people use
varying criteria in assigning MR. The experimental features are set up to be similar
in all relevant respects, save one. That respect is varied and it is found that MR
attributions vary with the varied criterial factor. Thus, on the issue of whether
people are compatibilists or incompatibilists, the findings indicate that whether the
issue is presented in an abstract or concrete manner makes a difference. Abstract
presentations bring out incompatibilist responses and concrete ones compatibilist
answers.

One supposition of invariance theories is that the core concept allows for a
unified theory of praise and blame— one that accounts for both. That is, the same
set of criteria for MR can be applied to morally good, morally bad, or morally
neutral behavior. Empirical evidence suggests, however, that there are at least five
asymmetries in MR attributions indicating that no single set of criteria for MR assignment functions in assessing the wrongness or rightness of an action or in determining whether it is to be blamed or praised.

Roughly, here are the asymmetries. First, the side-effect asymmetry shows that people vary their responsibility attributions depending upon whether a foreseen side-effect is morally bad or good. Responsibility attributions for foreseen morally good side-effects are reduced, while those for morally bad ones are not. Secondly, the emotion asymmetry effect shows the following pattern. If overwhelming emotion produces an action, it leads people to reduce the MR that they attribute to those performing morally bad actions, but not to those who have done morally good deeds. Third is the intention/action asymmetry. People are blamed almost as much for bad intentions as they are bad actions, but they are not given nearly as much praise for good intentions as they are for good actions. The moral ignorance asymmetry is a fourth sort of asymmetry revealed by recent empirical work. That an agent is morally ignorant leads people to think that she is less blameworthy, but it does not lead them to consider an agent less worthy of praise. Lastly, the fifth asymmetry concerns the role of negligence in the assignment of MR. It is generally agreed that people usually draw a line above which an agent is not considered to be negligent. Invariance theorists assume that line applies to both severe and mild harms. Being not sufficiently careful is one matter and the harmfulness of an action, for instance, is another. However, it turns out that people consider an agent to be responsible for severe harms, even if her action was only slightly negligent, but they do not allow that an agent is morally responsible for a mild harm unless she was extremely negligent.

Knobe and Doris (2010) sum up these findings in this way: “These five studies all used the same basic structure. People were presented with behaviors that differed in their moral status but seemed highly similar in every other respect. It was then shown that people ascribed a lot of praise or blame to one of the behaviors but not the other” (pp. 337-338). They conclude that though perhaps an invariantist theory can explain one or two of these asymmetries, it cannot explain all of them. Even so, an invariantist still might maintain that there are core features that are always operative in MR assignments. They can argue that moral agents are intentional agents that bring about certain behaviors whether morally good, bad, or indifferent. Attributions of MR correlate with the possession of these characteristics by moral agents.

Knobe and Doris (2010) concede that it is at least possible that judgments of MR are based on antecedent judgments about causality and intentionality. However, they maintain that there is good reason to think that these judgments themselves vary. The criteria for their application vary, and they vary with the goodness or badness of the behavior. The side-effect asymmetry brings this out since the attribution of intentionality to the agent is correlated with whether the foreseen side-effect is morally good or bad. Studies also indicate that attributions of causality vary with the goodness or badness of the actions that they bring about. These studies present cases of an agent who causes an accident while in the process of doing something morally bad, morally good, or neutral. Assessments of
the degree to which the agent is said to have caused the accident vary with whether
the action was morally good, bad, or neutral. People attribute more causal
influence to an agent bringing about an accident in the pursuit of something
morally evil than in the pursuit of something morally good or neutral.

Despite these findings, most social psychologists continue to distinguish
between *competence* in the assignment of causation and *performance*. The
competence mechanism is considered to be one that functions in a purely
descriptive fashion making use of relevant factual criteria for causal agency. It is
assumed that moral considerations play no role in the functioning of the
competence mechanism. The performance mechanism is one that determines
whether, in particular cases, the criteria for casual agency are instantiated. This
mechanism can be influenced by moral considerations that might distort or bias
judgments about causality. Thus, it is possible to interpret the above findings as
performance errors rather than displaying features of the way that the competence
mechanism works.

However, there are philosophers who maintain that there is, in some cases, a
tighter connection between causation and moral assessment. For instance, Knobe
and Fraser (2008) have presented a study in which actors do similar things—take
pens from an office supply station. But the administrative assistants who do this
are allowed to do so, while the professor who does so has been told that it is not
permissible. The office runs out of pens. Who caused the shortage? Respondents
indicate that it is the professor. Such studies seem to indicate that at least in some
situations moral assessments play a role in the competence mechanisms. But
Knobe and Doris (2010) argue that even if this is not the case, judgments about the
moral goodness or badness of an action may still always play an indirect role in
MR judgments by affecting causal judgments that, in turn, affect judgments of MR.
If so, even if the agent’s causal activities are always an object of assessment in
attributions of MR, that assessment can be influenced by moral factors in the
operation of the performance mechanism.

Finally, variance theorists point to a number of studies that indicate that
ordinary everyday assessments of MR vary with the relationship between the
person doing the assessing and the person being assessed. Philosophers who have
explored assessments of responsibility with thought experiments usually imagine
situations in which the parties involved are strangers; indeed, sometimes one or the
other is some sort of strange alien creature. These thought experiments often lead
to conclusions that support an invariantist theory. Thus, philosophers often neglect
most, and the most important, real-life situations in which MR attributions are
made. These are cases in which the parties know each other and are often involved
in relationships with each other. It turns out, however, that social psychologists
often use test stimuli that involve the interaction of strangers. Nevertheless, there
are a number of studies that indicate that people’s assessments of cases vary
widely with the relationships between the parties involved.10 And in these

---

10 I omit here the multitude of cross-cultural evidence for the variation of criteria for
assigning MR. Sommers (2012) is a good source for a lot of this evidence.
situations, attributions of responsibility vary with the relationships, for instance, whether the parties are friends, spouses, relatives, co-workers, or otherwise affiliated.

Let us return to the competence–performance distinction. Under the assumption of the variance hypothesis, people use different criteria to assess MR. Given these findings, two options open up. These criteria, including the moral goodness or badness of the action being assessed, are factors in the exercise of a person’s competence. Another hypothesis is that the various factors affect performance—the application of the criteria for assessing MR in certain sorts of situations, for instance. Both of these hypotheses are empirical ones that need to be tested.

Knobe and Doris (2010) emphasize that the psychological distinction between competence and performance mechanisms is not the same sort of distinction as the normatively moral one about correct or incorrect attributions of MR. Above I have distinguished (1) the descriptive question (what concept(s) are people using when they talk about MR?) from (2) the epistemically normative question (what phenomenon or phenomena, if any, are their conceptions tracking?). And I have distinguished both questions from (3) the morally normative question (what sort of response is appropriate for the tracked reality?). Knobe and Doris’s (2010) questions correspond to my (1) and (3). They call the family of views that contend that the criteria currently in use for assigning MR are more or less successful in correctly attributing MR “conservatism.” Revisionism, on the other hand, is the family of views that maintains that our current criteria are deeply flawed, and that to attain correct assignments these criteria must be altered. Suppose variantism captures current practice, making invariantism empirically implausible. Invariantists can adopt revisionism and argue that normative practice requires invariantism—but how much revising should be done? Knobe and Doris (2010) envision revision along a continuum and speculate that the need for revision will fall somewhere in the middle rather than at its extremes.

Waller, of course, is an advocate of eliminativism—what we might think of as total revision. But, as we have seen, Waller assumes an invariantist view about MR attribution insofar as what he has in mind is a response-invoking feature that is constituted either by libertarian or compatibilist accounts of FW. Nevertheless, there seems to be no reason why he might not embrace the evidence we have examined for the use of variable criteria in the assignment of MR. After all, these are still criteria for assessing MR. Indeed, he might well find the evidence for variance as further reason for arguing for elimination. The diversity itself, he might argue, is itself reason to suspect its ability to track a genuine phenomenon, let alone a phenomenon that then can generate morally correct responses in terms of fairness. Thus, he might well argue that elimination is superior to conservatism or revisionism, whether invariantism or variantism is correct.

However, this suggestion in fact outlines an agenda for Waller. For, as I have argued, if variantism is correct, W3, in either its W3L or W3C versions, is problematic. Consequently, Waller’s intended critique of the core concept of MR
seems to be a critique of, at best, a philosophical construct rather than the actual practice of ordinary moral agents.

Variantism also threatens W4. That premise asserts the unfairness of assignments of MR, but it is based on cases where the core concept is an invariantist one that connects MR to either libertarian or compatibilist FW. Waller has offered no argument establishing the unfairness of response practices that are based on the varying response-evoking features that seem to be operative in folk practice—those that reflect varying conceptions of that response feature. W4 might be justified but irrelevant to actual practice if that is best characterized by variantism. Even more problematic for Waller’s view is the evidence concerning the way people sometimes attribute MR. Rather than focusing on some response-evoking condition connected with crucial causal control (the MR-FW connection) as the basis for attributing MR and, thus, giving praise or blame, they appear to invert the process. Their assessments of moral goodness and badness, as well as their consequent offering of praise and blame, govern their judgments about the degree of critical causal control or the presence or absence of central factors relevant to the presence of causal control.

But even if we assume invariantism, W4 is problematic. Recall W4: The practices of praising and blaming and of rewarding and punishing can be fair only if they are responses to actions over which a moral agent has crucial casual control. In his critiques of both libertarian and compatibilist accounts of MR Waller employs two strategies. He appeals to a principle of fairness that I have tried to capture in W4, and in defense of that principle he constructs vivid stories of moral agents, their deeds, and their characters. He then asks the reader to make judgments about the fairness of blaming or praising them for their actions. Often he pairs agents who are successful and unsuccessful in their efforts. As I have mentioned, two ways of understanding Waller’s argumentative strategy come immediately to mind. First, the reader might understand that Waller is employing the usual philosophical method of cases. One tests one’s theory about a philosophical position—here an ethical principle—by examining imaginary cases. If the case fits the principle, then the principle gains confirmation. If not, then the principle might need to be revised or even withdrawn, or further thought might need to be given to the test case and its implications. Second, Waller’s method might be viewed from the explicit naturalist stance that he adopts. I suggested above that Waller’s vivid imaginary cases, along with his discussions of some real-world cases of crime and punishment, might be viewed as illustrations of naturalistically-based moral principles. This latter way of understanding Waller has the advantage of taking his oft invocation of naturalism seriously. It also has the advantage of understanding Waller in such a way as to avoid the problem to which the standard philosophical method of appeal to cases succumbs. This is the problem of an irresolvable battle of intuitions. However, although Waller frequently appeals to his naturalistic philosophical stance, he does not flesh it out enough to determine which interpretation of his argumentative strategy best fits what he is doing. Waller’s naturalism commits him substantively to denying supernatural agents, miracles, and other sorts of non-natural causal factors such as
agent causality. His use of scientific findings in his rejection of MR and the development of his own positive position also commits him substantively to some sort of view that connects nature, norms, and values in a seamless way. But the nature of that connection remains implicit and undeveloped. Thus, it remains unclear how to assess W4 even under an invariantist assumption.

Scientifically-based naturalistic theorists will reject the intuition-based method of cases and develop conclusions about moral values and norms using the best current scientific findings and theories. Thus, despite the fact that his appeal to naturalism provides a prima facie indication that he is taking the right approach to the issue of MR, Waller’s actual argumentative strategy remains ambiguous enough to make the scientific naturalist pause. Moreover, intuition-based arguments for unfairness attributions of MR will undoubtedly fail to convince either libertarian or compatibilist proponents of MR, or both. They surely will be unpersuasive to scientific naturalists. On the other hand, his normative arguments based on well-established facts of the causal relations in the world, including those psychological findings about human moral agents that Waller uses to support his account of FW and TCR, appear to be undeveloped or, more seriously, beg the question against proponents of compatibilist MR.

Waller bases his charge in W5 that the MRS system is unfair on the application of W5L, the libertarian MR-FW connection. So it begs the question against advocates of the compatibilist MR-FW link. The compatibilist proponent of MR maintains that if it is the case that genuine moral agents exist, as Waller himself argues, then causal agency exists at various levels of complexity, including the psychological. Agents with cognitive capacities such as a sense of self-efficacy and of an internal locus of control genuinely accomplish something by means of these capacities or fail to do so because of failures of these capacities. In such instances of success or failure (or degrees thereof), success or failure is properly attributed to the operation of these factors, not to either currently operating lower-level capacities or external environmental factors, nor to past environmental or lower-level factors that played a role in the acquisition of these factors. Waller confounds the failure of moral capacities to function properly with the failure of the non-moral capacities that are the pre-conditions for the presence of any MR.

But Waller might object that a moral agent’s cognitive and motivational capacities all depend, for their current operation, both on lower-level internal capacities and sustaining external environmental factors. In doing so, Waller might understand the relationship between lower-level capacities and the higher-level ones—those constitutive of moral agency—to be one of supervenience. Supervenience relations come in various forms. The weaker versions postulate a correlation between lower-level and higher-level factors. A stronger version understands the relationship to be one of determination.11 Weaker versions are not

---

11 The determination relation takes two alternative forms: (1) a constitutive relation such that lower-level factors are constitutive elements of the higher-level factor or (2) a causal relation where the operations of lower-level factors causally determine the operations of higher-level factors.
sufficient to establish the dependency relation. However, the stronger
determination relation is too strong because, on that account, the causal capacities
of the higher level become epiphenomenal. Thus, that sort of dependency
relationship between lower-level capacities and those of moral agency renders the
causal efficacy of a moral agent’s capacities epiphenomenal. Interpreting Waller’s
view of the dependency relation in this fashion explains why he offers
justifications only for W5L. It also explains why, in addressing W5C, he turns to
lower-level factors to determine whether attributions of MR are ever aptly applied
and argues that they are not since variations in these lower-level activities imply
variations in higher-level activities. On this understanding of Waller’s view,
distinct levels of causal capacities seem to collapse.

Consequently, Waller’s naturalism turns out to be a physicalistic naturalism.
Though it is non-reductive ontologically insofar as it allows that humans have
distinct capacities operating at different ontological levels, it is explanatorily
reductive since it attributes causal efficacy only to lower-level causal factors,
perhaps only to the fundamental entities discovered or to be discovered by physics
and to external environmental factors that also have their causal powers in only an
epiphenomenal fashion. However, physicalistic naturalism appears to be
inconsistent with Waller’s claim that humans are genuine moral agents who really
perform good and evil moral actions and fashion their character and themselves in
morally good and bad ways, producing genuine moral goods and evils to which
these same moral agents appropriately respond.

If this analysis is correct, then Waller’s view also suffers from the fact that it
requires a deflationary account of the psychological findings that he uses to
support his account of compatibilist freedom, those concerning self-efficacy, locus
of control, and cognitive pursuit. The randomized, controlled studies upon which
these findings are often based provide the justification for claims that humans have
cognitively-based capacities that exert genuine causal influence on their behaviors,
their character, and the selves that they become, as well as on others with whom
they engage. A substantive and methodological scientific naturalist should use
these results in fashioning her position rather than make use of a principle like
supervenience, for that principle is far from being an established scientific one and
is, to a large extent, incompatible with what even the physical sciences are telling
us about causal agency in this world (Figdor, 2010; Morrison, 2012; Potochnik &

In summary, I have argued that empirically-based findings, some of which
Waller himself appeals to, render premises W3, W4, and W5 of Waller’s negative
argument problematic. If I am correct in my critique, Waller’s argument against
the MRS is unsound and its conclusion against the MRS fails. Of course, I may
have miscast Waller’s argument and my challenges to W3, W4, and W5 may be
shown also to fail. But, as things stand, I contend that Waller has not made his case
against the MRS.

But what can we say about his positive project? I now turn to that proposal
and an assessment of it.
The Successor: Take-Charge Responsibility (TCR) System

In the course of his argument against the MRS, Waller informally develops his own positive view of a successor to the MRS, a TCR system.12 I reconstruct it as follows:

**The Argument for a TCR System**

(1) Humans are genuine agents with causally efficacious cognitive and motivational capacities that bring about actions that affect themselves and others. (Agency hypothesis)

(2) This agency is operative in the moral realm. Thus there are genuinely morally good and bad actions, characters, and consequences. (Moral agency hypothesis)

(3) In exercising their moral agency, humans have the capacity to control what they are doing. Thus they have a form of compatibilist freedom. (Compatibilist FW hypothesis)

(4) That freedom enables them to take responsibility for what they do, become, and bring about. That is, they possess TCR. (Take-charge responsibility hypothesis)

(5) As moral agents they respond to their own and others’ good and bad moral actions, the characters that develop on the bases of these actions, and the effects that their actions and characters have on others. Consequently, they rightfully praise and admonish each other and express gratitude, regret, and the entire range of moral responses. (Appropriate response hypothesis)

(6) But these responses must be distinguished from those of the MRS; for instance, we must distinguish attributions of blame fault from those of character fault. (The distinction hypothesis)

(7) The TCR system can be adopted and developed. (Psychological realism hypothesis)

(8) The TCR system is a morally praiseworthy and beneficial responsibility system. (Moral appropriateness hypothesis)

---

12 Waller does not speak in terms of a “take-control responsibility system.” He proposes TCR as an alternative to MR. However, it seems clear that he has in mind individuals, institutions, and societies that exercise TCR.
(9) The MRS is a morally deficient and destructive responsibility system and thus ought to be abandoned. (Conclusion of the negative arguments against the MRS)

(10) All things considered, TCR ought to be adopted and developed. (Follows from (7), (8), and (9))

Assessing Waller’s Positive Proposal

Space does not permit assessing this proposal in any detail. However, I think that there is substantial empirical and theoretical scientific support for the first five premises of Waller’s argument. Waller himself appeals to some of this support. Humans are moral agents that exercise their moral agency through the capacities that compose compatibilist freedom. Given those capacities humans can and do take charge of what they do.\(^\text{13}\) And they can and do respond appropriately to other moral agents and their actions. A scientific naturalist critic of Waller can find herself in agreement with Waller on these important premises.

Of the remaining premises, (9) is the conclusion of the negative argument upon which we have already commented in detail. Premise (6) maintains that MR and TCR are distinct. Premise (7) maintains that since TCR is a genuine phenomenon, a TCR system can be implemented and developed. Premise (8) claims that it ought to be.

Premise (6) asserts that the moral responses evoked in the MRS and TCR systems are distinct. I must admit that I am unable to discern a fundamental difference between the TCR and MRS systems. One might conjecture that the difference resides in the difference between a retributive and consequentialist sense of the fairness of blaming. However, Waller rejects consequentialist justifications of retributive justice (Chapter 8). Another possibility is that TCR offers a much better moral framework for human interactions. It suggests practices that improve individual and collective moral agency and their results. Whether that is so or not, this suggestion seems to give the same functional role to TCR and MRS, and to that extent it does not distinguish them. Moreover, as we have seen, Waller is not seeking to improve the MRS but to replace it. Even if it is the case, as Waller thinks it is, that TCR is a huge improvement over MRS, it accomplishes that feat not by performing the same function as MRS but by being something else. I am not able to discern what that something else is.

On the other hand, Waller also tells us that the TCR system is a lot like MRS. We can see that because the first five premises of the argument for the TCR system present elements of TCR that seem to be identical with those of MR. To that extent, even given the difficulties in pinning down the features of TCR that distinguish it

\(^{13}\) Waller’s account of TCR finds its basis in such empirically established psychological features as (1) having a sense of internal locus of control, (2) a sense of self-efficacy, (3) chronic cognizing, and (4) cognitive fortitude. The existence and nature of these features (aspects of compatibilist FW) make for successful agency. Thus, Waller forges a link between TCR and compatibilist FW.
from MR, the assertions about the psychological realism of TCR and its moral worth in Premises (7) and (8) receive support from the bases for Premises (1)–(5).

However, the way in which Waller develops his case for the psychological realism of TCR and the possibilities for its adoption, as well as the bases for its moral worth, raise difficulties for both Premises (7) and (8). This is so because of the way that Waller seems to understand the emotional bases for the rejection of the MRS and because of his critique of attempted philosophical justifications of the MRS as mere rationalizations. I will expand some on this problem.

**Replacing the MRS**

Chapter 15 describes a world without MR. In some very important ways the world is very much the same as the one in which we now live:

> We would still make free choices, moral judgments, and sincere apologies, still feel affection and gratitude, still recognize our faults and strive (more effectively) to correct them. We could retain and strengthen our self-respect and our sense of individual worth. . . (p. 278)

Nor would all responsibility disappear, for TCR will remain and be enhanced. Moreover, the process of abolishing the MRS will make apparent the evils of the strike-back emotion response and uncover the blindness to the historical antecedents (both genetic and learning) of character and actions that trap advocates of the MRS. In addition, it will reveal the futility of retaining the MRS despite the recognition that it is illusory and reveal the inadequacies of neglecting the differences in agential capacity of those who have attained the so-called MR plateau. And, very importantly, attention will be shifted from “morally irresponsible” individuals to systematic failures that prevent individuals attaining moral successes and that foster their moral failures.

For those who might find these predictions overly optimistic and who might want something more concrete and tangible, Waller points to practices in manufacturing, aviation, and medicine that look to the big picture and to how systems operate well and poorly rather than focusing on the individuals within the system. And he points to the restorative justice movement as an alternative to a failing criminal system. Waller argues that this shift in focus need not let the criminals off scot-free, but it does require moving from blame fault to character fault and attempts to address the latter.

But is genuine and complete replacement really possible? Perhaps the MRS is too deeply rooted in an irrational, but ineliminable, strike-back emotion, too ensconced in a very long cultural history and individual learning histories, too susceptible to the wiles of rationalization that appear to infest our attempts to determine what is the morally right thing to do. Indeed, it may be especially hard for philosophers to dislodge their attachment since they are, to a large extent, privileged actors who are blessed with the virtue of those who persist in deliberation and possess a strong sense of internal locus of control and of self-
efficacy. In his final Chapter, Chapter 16, Waller argues vigorously for the real possibility of elimination, despite all these apparently invincible impediments.

Waller compares the philosophical efforts of the advocates of the MRS to a Lakatosian degenerating research program. The signs of failure are marked in particular by the plethora of different attempts by both libertarians and compatibilists to assure its survival despite all the evidence against it. He summarizes that evidence—evidence that he has detailed in the earlier chapters:

Yet for all our [philosophers'] inclination and desire to believe in moral responsibility we also recognize that moral responsibility and the claims on which it rests are threatened on every side. Neither luck nor talents average out. Our capacity to continue deliberating is a product of our history (Cacioppo et al. 1996). Our deliberations function more to prop up beliefs already intuitively established than to rationally examine them (Haidt 2001). Our abilities to choose are limited by our locus of control (Rotter 1966) and sense of self-efficacy (Bandura 1997), which are a product of our fortunate or unfortunate conditioning histories and of which we are only dimly aware. Situations control our choices more than we ever imagined (Doris 2002). Our basic moral stances are given by our history, genetic and cultural, not our choice. Our willpower is epiphenomenal (Wegner 2002). Decisions are made nonconsciously (Libet et al. 1983). (p. 307)

Yet philosophers cling to the MRS, even though some like Smilansky (2000) recognize it as an illusion. Waller’s diagnosis of why most philosophers still cling to it is that the MRS system is deeply rooted in an irrational strike-back emotion that is fortified by common sense and legal systems. Given also that philosophers are cognitive misers—able and willing to consider issues in depth and detail—as well as being gifted with a strong sense of self-efficacy and internal locus of control, they find it easy to credit themselves for their accomplishments. These all reinforce the view that FW requires MR.

Our situation seems to be this: Though the scientific evidence indicates that we possess compatibilist freedom, it also makes clear that we lack ultimate control of what we do. And careful attention to the fact that we lack ultimate control highlights the unfairness of the practices of praising and blaming that are central to the MRS. Yet, at the same time, our irrational strike-back emotion fortified by our social and legal environment leads us to rationalize the fairness of these practices. So Waller asks himself:

Even if we acknowledge that moral responsibility is a harmful doctrine, that it is battered by scientific advances in our understanding of human behavior, that it has no rational justification, and that it cannot withstand close scrutiny, even if we acknowledge all that, can we actually consider renouncing moral responsibility? (p. 309)

By now we know Waller’s answer. Not only can we consider renouncing it, we are able to do so. Not only are the philosophical barriers penetrable, they have fallen apart, and there is every reason to think that the psychological barriers can
be overcome. Yet the naturalistic critic will find herself with some doubts, doubts raised by Waller’s own views about both our psychological and philosophical capacities.

Waller subscribes to the claim of social psychologists Jonathan Haidt and Fredrik Bjorklund (2008) that moral reasoning is often an exercise in rationalizing one’s prior emotionally-based intuitions about what is right and wrong. He has used this view to diagnose why it is that we cling to the notion of MR despite all the counter-evidence. The reason is that we are held in the grip of the strike-back emotion. Nevertheless, Waller claims that Haidt and Bjorklund’s empirically-based theory of what counts as belonging to the moral realm might provide us with a way to understand how we might break the grip of the strike-back emotion that, on Waller’s view, constitutes the core of the MRS. Haidt and Bjorklund have argued that the moral realm includes five distinct areas, each supported by an intuitive emotional response. These concern fairness/reciprocity, harm/care, authority/respect, in group/out group, and purity/sanctity. Waller argues that the emotionally-based source for our intuitive sense of fairness has offered, and can continue to offer, a way to offset the influence of the strike-back reaction. There is historical evidence that the latter has been modified and pared back. He also contends that knowledge of the sources of our actions, decisions, and character can further the employment of the fairness intuition. But it is not at all clear whether these two reactions—that of the strike-back emotion and the fairness emotion—are, in fact, distinct. Their functions both seem to be to respond in an appropriate way to the moral or immoral actions of another. In answering whether we have the psychological capacity to free ourselves from the strike-back emotion, Waller proposes that we use other emotional resources to counter it. But the emotional resource that he calls upon is not clearly distinct from the one it is called upon to counter.

Moreover, there are problems on the philosophical side concerning the issue of justification. Waller’s strategy is in tension with his acceptance of the thesis that moral decision-making is rationalization. Haidt and Bjorklund (2008) have argued that when people are confronted with challenges to their moral views they defend them. When these responses are shown to be unsatisfactory, they cling to their original view, providing no further justification for it. Their response to challenge is: “It just is correct and that’s it!” Haidt and Bjorklund have characterized this sort of response to moral challenge as moral dumbfounding, and they have used the phenomenon of moral dumbfounding to argue that moral reasoning is not an attempt to find justifying reasons for a moral position, but a process of rationalization. That process is a way of defending basic, irrational, emotional responses. On this account there is no role for a process of moral reasoning that provides justificatory reasons for moral actions.

But Waller maintains that dumbfounding might, in cases of the confrontation of intuitions about retaliation and fairness, lead to genuine moral reasoning, moral

---

14 Waller seems to use “intuition” and “emotional response” interchangeably in this discussion.
reasoning that takes the side of fairness. He goes on to present some evidence of this happening. He compares cultures (social democratic ones) that control retributive instincts with those (neo-liberal ones) that seem to celebrate them (pp. 314-315). The cultural phenomenon parallels the individual phenomenon of impulse control. Finally, Waller points to significant changes in moral practices: the rejection of religious intolerance, the acceptance of homosexuality, the acceptance of interracial marriage, and the acceptance of once morally unacceptable sexual practices. Indeed, Waller adverts to the trend of removing types of practices from the moral realm, thus the trend in Europe and North America to take practices associated with purity and sanctity as matter of non-moral preference. All this indicates the actuality and possibility of controlling, if not eliminating, the strike-back attitude. Thus, Waller’s use of Haidt’s account of moral reasoning pulls him in two directions. He uses Haidt’s rationalization account of moral reasoning to problematize philosophical attempts to justify the MRS, yet in making a case for its abandonment he claims that the fairness reaction can be used to motivate and justify the abandonment of the strike-back reaction associated with the MRS. But how, on Waller’s account, are such changes in moral practices any more justified than the ones that they replace, given Waller’s earlier appeals to Haidt’s rationalization hypothesis about moral reasoning? Given a distinction between the emotionally-based fairness intuition and the strike-back emotion—a problematic one, as we have seen—how can the former shed its role as a source for rationalization in moral reasoning and take on the quite different role of providing reasons that justify moral reasoning and the practices that such reasoning might prescribe? According to Waller, moral reasoning cannot justify the MRS because it is mere rationalization of an emotional response, but it can be used to justify TCR even though it, too, is based on an emotional response. Waller needs to explain why moral reasoning is justificatory in the latter case but merely rationalization in the former.

Thus, though a scientific naturalistic critic finds herself in agreement with Premises (7) and (8) of Waller’s positive argument, she has doubts about whether Waller has given us the tools to establish either that TCR is psychologically realistic or that TCR is morally justifiable.

Conclusion

In this critical review I have attempted to play the role of a friendly critic of Waller’s ambitious proposal to abolish the MR system and replace it with a TCR system. In doing so I have adopted a scientifically-based naturalistic approach, one, I believe, that is consonant with his own. I have argued that (1) his negative argument for replacing the MRS falters, (2) I have been unable to distinguish his proposed replacement TCR system from a reformed MR system, and (3) Waller has not made the case that TCR is either psychologically realistic or morally justifiable. Nevertheless, the issues that these critiques raise are, I believe, ones that further scientific theoretical and empirical work can clarify and resolve. Thus, in conclusion, I want to emphasize one of the aspects of Waller’s work that makes
it important for current philosophical work—its commitment to solving philosophical problems using the best available findings of the relevant sciences. This, along with his passion for justice and fairness—a passion that exudes from every page of his book—makes it a work very worthy of the attention of both philosophers and behavioral scientists.

References