



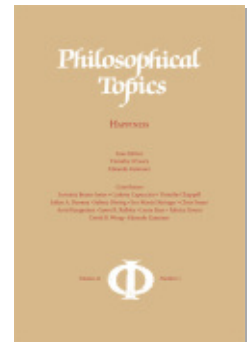
PROJECT MUSE®

Value Pluralism, Intuitions, and Reflective Equilibrium

Lisa Tessman

Philosophical Topics, Volume 41, Number 2, Fall 2013, pp. 175-201
(Article)

Published by University of Arkansas Press



➔ For additional information about this article

<http://muse.jhu.edu/journals/pht/summary/v041/41.2.tessman.html>

Value Pluralism, Intuitions, and Reflective Equilibrium

Lisa Tessman
Binghamton University, SUNY

ABSTRACT. A constructivist approach to ethics must include some process—such as John Rawls’s reflective equilibrium—for moving from initial evaluative judgments to those that one can affirm. Margaret Urban Walker’s feminist version of reflective equilibrium incorporates what she calls “transparency testing” to weed out pernicious, ideologically shaped intuitions. However, in light of empirical work on the plurality of values and on the cognitive processes through which people arrive at moral judgments (i.e., an automatic, intuitive process and/or a controlled reasoning process), I raise one concern: some moral requirements can only be grasped intuitively and should not have to be affirmed from the perspective of other confidently held values. For Harry Frankfurt, the “requirements of love” are one such example; failing to fulfill these requirements is, for someone who loves, unthinkable; one transgresses the associated values merely by *considering* sacrificing them. I suggest—citing empirical work on “sacred values” (such as the work of Philip Tetlock)—that to subject these requirements to transparency testing would be to transgress them by having “one thought too many” (as in the work of Bernard Williams). One’s confidence in these values and the authority of these values depend on an automatic process. I consider the risks, and the necessity, of embracing both intuitive and reasoning processes for affirming the authority of a plurality of moral values.

The starting point of this article is a particular kind of normative experience, the experience of apprehending a situation as *requiring* a certain response, or, put differently, the experience of reaching an evaluative judgment that may be expressed, most simply, by the words “I *must*.” This experience can vary in many ways, and I believe that the resultant plurality should be reflected in any account of morality that takes moral requirements to be constructed out of normative or evaluative experience. There is plurality in *what* humans experience as valuable and as requiring a response that preserves or protects the value. Some values and their accompanying requirements become moralized—imbued with moral authority—through a further plurality of social processes. The diversity of people engaged in these social processes serves as an additional source of plurality. And, as I will be emphasizing in this article, there is also plurality in *how* humans value, or make evaluative judgments; that is, in the cognitive processes involved in valuing or in arriving at moral or other evaluative judgments, as well as in making second-order judgments that can affirm (or reject) the initial judgments.

My conviction that moral values are plural is tied to my conception of what morality is. I take morality(/ies) to be something that humans, as an ultrasocial species, have naturally constructed, and that is necessary in order for us to live good enough lives together. Moral values—and their associated requirements, norms, practices, traits, structures, and so on—all count as moral because they each in some way serve to connect humans to one another, to enable or regulate at least one of the many forms of cooperation in which we engage, and to balance self-regarding and other-regarding concerns in doing so; their counting as moral—or our recognizing them as moral—depends upon their being imbued, through social processes, with a special kind of authority, and people’s confidence in the authority of morality makes it possible for morality to serve the function that it does. Moral value pluralism reflects the fact that there is a great range of values (and requirements, norms, practices, and so on) that, given this conception of morality, can count as moral; there is no single value that can do all that morality does for humans.¹

The literature on moral value pluralism has focused largely on what happens when different kinds of values conflict—for instance, in a dilemma—and one must choose between incommensurables, or else has focused on the question of whether value pluralism entails relativism, and if so, what kind of relativism.² These are important discussions about the implications of pluralism, but here I focus on a different concern: constructivist accounts of morality tend to exhibit a monism about what sort of a process of construction yields moral authority. My claim is that although this process—a critical reasoning process—can successfully imbue some kinds of values with moral authority, it is the wrong sort of process for generating the moral authority of other kinds of values and their associated requirements.

I am going to focus on the kind of experience—an experience of valuing and of judging oneself to be required to respond in a certain way—that constructivist accounts of morality may not countenance—namely, the experience of encountering what one takes to be a nonnegotiable moral requirement, or to put it even more strongly, of judging that the alternatives to a particular action are *unthinkable*, or that the value that some requirement preserves is *sacred*. The experience of what one might think of as the requirements of love³ often take this shape: if one's loved one is in danger, it is unthinkable to abandon him/her; the requirement to protect him/her is nonnegotiable. But unthinkability can also be involved in cases that are not love relationships with particular others. For instance, for many people, it is unthinkable to kill another human being, at least up close. Of course, sometimes, in a tragic conflict, one can be forced to think—and to do—the unthinkable, so “unthinkable” does not mean, literally, “unable to be thought.” Rather, if the sacrifice of some value is unthinkable, what it means is that this sacrifice is “unable to be thought or considered without thereby transgressing or violating the value.” In these cases, the mere contemplation of sacrificing the value is transgressive, even if the sacrifice never takes place. Empirical research shows that among the plurality of values that people experience, some are experienced as sacred, and that among the plurality of associated requirements that people grasp, some are experienced as nonnegotiable—and their alternatives as unthinkable.

The problem that I am interested in arises when one tries to find a place for these sacralized values within a constructivist account of morality; namely, an account that, in Sharon Street's words, takes value to be “a construction of the attitude of valuing.”⁴ According to Street, “*Constructivist views in ethics* understand the correctness or incorrectness of some (specified) set of normative judgments as a question of whether those judgments withstand some (specified) procedure of scrutiny from the standpoint of some (specified) set of further normative judgments.”⁵ In what she calls a restricted form of constructivism, the grounding judgments (the judgments from the standpoint of which one scrutinizes other judgments) are given or presupposed, but I will be examining what she calls *metaethical constructivism*: “According to *metaethical constructivism*, the fact that *X* is a reason to *Y* for *A* is constituted by the fact that the judgment that *X* is a reason to *Y* (for *A*) withstands scrutiny from the standpoint of *A*'s other judgments about reasons.”⁶ In metaethical constructivism, every judgment must take its turn being scrutinized from the standpoint of one's other judgments in what I will call the “process of construction.” The process of construction is what allows one to move from initial, often intuitive, valuations to those that are to be affirmed and imbued with authority. Many constructivists (including Street) have taken this process to be a process of reflective equilibrium (borrowing primarily from Rawls). I am also assuming that it is a Neurathian process; namely, a process that is like what sailors must do to repair their ship while at sea: there is no foundation—nowhere off the ship—on which to stand while assessing and repairing the planks of the ship or, analogously, one's values.⁷

Where I disagree with other constructivists is in my belief that there are cases in which neither rational reflection nor equilibrium are *necessary* components of the construction of authoritative reasons. That is, they are not necessary components of a procedure that can lead to the confident endorsement of a moral judgment. Discussions of the process of construction have not, I believe, reckoned sufficiently with value pluralism and its implications. My concern is that the plurality of what is valued, and of how evaluative judging takes place, is not reflected in a method that is singular in its prescribed route to moral authority, and which takes moral authority to always depend upon justification. In particular, I believe that those who adopt a version of reflective equilibrium as the process of construction do not have in mind sacred values and what it takes to preserve rather than destroy them. I thus raise a problem with the method of reflective equilibrium, suggesting that among the plurality of values worth preserving there are some values—sacred values—that would be violated by being subjected to the method of reflective equilibrium. My focus will be on Margaret Urban Walker’s feminist version of reflective equilibrium, which seems to be motivated largely by a particular, and in my view valid, concern about intuitions—namely, that they have been ideologically shaped and tend to reinforce relations of dominance and subordination. I aim to show that we are in a bit of a fix: subjecting every intuitive judgment to reasoned reflection can destroy some of what we value—in fact, it can destroy some of the best of what we value—but exempting intuitive judgments from such scrutiny is dangerous, and this danger is particularly evident from a feminist perspective.

II

I will return to a discussion of the process of construction, but first I will present the two empirical findings that I will rely on in making my point: (1) people make most of their moral judgments intuitively and automatically, instead of through a controlled reasoning process, and most changes in moral judgments take place through a social process in which people influence each other’s judgments in affect-laden ways; and (2) people experience some values as sacred, such that they find it to be forbidden to even consider sacrificing these values.

The claim that we make most of our moral or (more broadly) evaluative judgments intuitively is well supported by research in cognitive and social psychology. Cognitive psychologists widely recognize two psychological systems for cognitive processing. There is “System 1,” or the intuitive system, which “operates automatically and quickly, with little or no effort and no sense of voluntary control,” and “System 2,” or the reasoning system, which “allocates attention to the effortful mental activities that demand it.”⁸ System 1 is unconscious, associative, automatic rather than controlled, and fast. In contrast, System 2 is conscious, inferential,

controlled, and relatively slow.⁹ The operations of each system engage a number of different brain regions, and the two systems can be brought into conflict.¹⁰

A variety of scientific methods have been used to investigate specifically *moral* cognition, yielding a dual-process model of moral judgment that recognizes that “moral judgment is the product of both intuitive and rational psychological processes.”¹¹ Importantly, though, the research shows that *most* moral judgments are arrived at through the affect-driven, automatic, intuitive process; typically, an automatic process produces the judgment, and controlled reasoning takes place after the fact (*post hoc*), to produce justification for the judgment, rather than to produce the judgment itself.¹² However, people who form moral judgments in this way are not aware that they are doing so; they tend to believe they have reasoned their way to the judgment.¹³

Jonathan Haidt has developed a dual-process model of moral judgment that he calls the social intuitionist model, a model that is meant to replace the now discredited rationalist model of moral judgment. According to the rationalist model, one reaches one’s judgment as a solo act, performed through the reasoning process. In contrast, in the social intuitionist model, when one person gives justifying reasons or arguments (which were formed *post hoc*; i.e., after the judgment was made) in support of a moral judgment, those reasons can affect *other* people, primarily by giving rise to intuitions in them. A person’s *own* reasoning is almost always “motivated” or biased in favor of supporting her/his own prior intuitive judgments, so private reasoning rarely brings about a change in one’s own moral judgments.¹⁴ However, another person’s reasoning—supporting an opposed judgment—is much more likely to change one’s judgments. People are also often affected by others’ judgments even when supporting reasons for the judgment are not supplied: “Because people are highly attuned to the emergence of group norms, the [social intuitionist] model proposes that the mere fact that friends, allies, and acquaintances have made a moral judgment exerts a direct influence on others, even if no reasoned persuasion is used.”¹⁵

Intuitive moral judgments differ from reasoned moral judgments in that they are not based on any *justifying reasons*. If making inferences (as distinguished from, say, unconscious associations) is essential to justification, and the intuitive process involves no inferences, then this process does not yield justifying reasons. In rationalist moral theories, the absence of justifying reasons is sufficient to warrant the disqualification of a judgment as normative or prescriptive. However, it is common to *experience* intuitive moral judgments as normative or prescriptive regardless of the fact that they come without justifying reasons.

I turn now to the second empirical finding on which I will rely, which comes from Philip Tetlock’s work on what he calls *sacred values*. Judgments about sacred values are one kind of moral judgment that is produced through the intuitive process described above. Tetlock defines a sacred value as “any value that a moral community implicitly or explicitly treats as possessing infinite or transcendental significance that precludes comparisons, trade-offs, or indeed any other mingling

with bounded or secular values.”¹⁶ When one sacralizes the value of something or someone, one also takes its sacrifice to be unthinkable—not just forbidden, but unthinkable, which means that any *consideration* of sacrifice is inappropriate. Tetlock’s research documents what results from this tendency to sacralize certain values.

In Tetlock’s and his colleagues’ experiments, subjects are presented with narratives about (fictional) decision makers who contemplate trading off a sacred value, even if they ultimately decide not to. The scenario might involve a decision to sacrifice one sacred value for another sacred value—which is called a “tragic trade-off”—or might involve the possibility of trading off a sacred value for a nonsacred value—referred to as a “taboo trade-off.” Experimenters look at what subjects’ reactions are to these stories, and in particular, at whether subjects react with moral outrage. For example, in a narrative about a hospital administrator who must decide whether to spend funds to save the life of a child or to use the same funds “for other hospital needs,” if subjects are told that the hospital administrator decides “after much time, thought, and contemplation” to save the child’s life, then the subjects express intense moral outrage about the administrator, but they do not if they are told that the administrator is very quick to make the decision to save the child’s life.¹⁷ In other words, thinking about the unthinkable (in this case, sacrificing a particular child’s life) is treated as a moral transgression, and the more one thinks, the worse it is: “Even when the hospital administrator ultimately affirmed life over money, his social identity was tarnished to the degree that observers believed that he lingered over that decision. It was as though participants reasoned ‘anyone who thinks that long about the dollar value of a child’s life is morally suspect.’”¹⁸ On the other hand, if the narrative is altered so that the hospital administrator must choose to either save the life of one child or save the life of another child, thus leading subjects to treat the situation as requiring a tragic (rather than taboo) trade-off, then they praise the administrator for spending *more* time deliberating; when the sacrifice of a sacred value is inevitable, longer deliberation signals a deeper desire to prevent this inevitable sacrifice.¹⁹

The subjects in these experiments treat the life of the named, particular child as a sacred value—the sacrifice of which is unthinkable. Sacred values exhibit what Joseph Raz (1986) calls “constitutive incommensurability.”²⁰ For Raz, values are constitutively incommensurable when any consideration of exchange—even if one does not in fact choose to exchange—undermines one of the values by revealing that one did not take it to be incomparable. Typically, personal relationships are valued in such a way that to consider exchanging them undermines them. As Tetlock et al. put it in their development of Raz’s concept:

Our commitments to other people require us to deny that we can compare certain things—in particular, things of finite value with things that we are normatively obligated to treat as infinitely important. To transgress this boundary, to attach a monetary value to one’s friendships, children, or loyalty to one’s country, is to disqualify oneself from

the accompanying social roles. Constitutive incommensurability can thus be said to exist whenever comparing values subverts one of the values (the putatively infinitely significant value) in the trade-off calculation . . . To compare is to destroy.²¹

I take Tetlock's and his colleagues' research to show that, psychologically, people take some things to be appropriately valued only when moral judgments about them are made intuitively, as reasoning about their value entails transgressing them. Any search for justification of an action aimed at preserving a sacred value is problematic if it suggests that the value in question would be abandoned if justification were to not be found, since abandoning a value is exactly what must remain unthinkable if the value has been sacralized. Reasoning that consists of weighing costs and benefits is especially problematic because this sort of reasoning implies the comparability and fungibility (rather than constitutive incommensurability) of values. The move from intuition to reasoning is a betrayal of values that are in part constituted by their guaranteed insulation from the negotiations that take place through conscious reasoning. People's psychological responses tend to reveal that they hold that when values have been sacralized such that they are constitutively incommensurable with nonsacralized values, one commits a moral wrongdoing whenever one engages in reasoning about the sacred value.

A similar idea was captured by Bernard Williams in a much-quoted passage. Williams writes about a hypothetical rescue situation in which the rescuer reasons that it is justifiable for him to preferentially rescue his wife rather than impartially decide whether to rescue his wife or a stranger. Responding to arguments that intend to demonstrate "that moral principle can legitimate his preference, yielding the conclusion that in situations of this kind it is at least all right (morally permissible) to save one's wife," Williams remarks:

This construction provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife.²²

In cases where, for instance, a loved one is in danger, one thinks one thought too many just by engaging the reasoning system to arrive at the judgment that it is permissible to give preference to one's loved one, instead of engaging the intuitive system that would automatically put one's loved one first. It is not that the rescuer in Williams's case wronged his wife by making the wrong decision; rather, his wife must judge that he did something morally wrong by thinking his way to the best decision in such a way that degraded the value (his relationship with his wife) about which he did his thinking.²³

The notion of a sacred value illuminates Williams's case of the rescuer. What the rescuer's wife might have hoped is that her value to him was sacred, that the possibility of not saving her first was, literally, unthinkable for him. When

he reasons about it instead of responding intuitively he reveals that she was not sacred to him after all. If Tetlock's subjects were to hear the story of the rescuer, they would presumably feel that he, like the hospital administrator, had committed a moral transgression just by virtue of having had "one thought too many."

III

The empirical findings that I have just presented suggest that intuitive processing plays an irreplaceable role in human morality: not only do we make most of our judgments intuitively, but there are some values that people have sacralized and thus marked as infinitely significant, their sacrifice unthinkable; upholding these values can only be accomplished intuitively, for reasoning ruins them and disqualifies the reasoner from the relationships whose core they form. I now turn to what motivates reasoned scrutiny of every intuition: our intuitive judgments often go wrong. Constructivists take scrutiny of all of one's moral judgments to be essential to the process of construction, precisely because we cannot trust all of our judgments, perhaps especially our intuitive judgments. If I am suggesting that sacred values must be protected from even the thought of their sacrifice, then scrutiny of judgments about sacred values will be prohibited. On the other hand, this scrutiny seems to be required if our intuitive judgments can and often do go wrong. Instead of considering all of the ways in which intuitive judgments could go wrong, I will focus on one that feminists have emphasized: our moral intuitions are likely to have been ideologically shaped. Some of our intuitive moral judgments are systematically and insidiously mistaken, such as those that exhibit implicit bias and that, if uncorrected, may lead us to perform racist or sexist acts. It is concerns like these that turn various normative theorists, including some feminist ethicists, to adopt some version of the method of reflective equilibrium.

The method of reflective equilibrium, as proffered by John Rawls and taken up by others, includes the following features: it takes judgments at all levels of generality (from particular judgments to more general principles and other convictions) to be relevant inputs; it involves reasoned reflection; it aims at the sort of equilibrium that is achieved when one's particular judgments and one's other more general principles and convictions all form a coherent set and thus stop exerting pressure to revise the set in opposing directions; and it takes initial judgments, principles, and theories to all be subject to adjustment in order to achieve this equilibrium—that is, the adjustment is mutual rather than unidirectional. In reflective equilibrium one tries to find principles and general theories that match and can explain one's particular intuitive judgments. If there is no match, one considers, in turn, how to adjust the principles so that they better fit one's most firm judgments, and one considers how to sort through one's judgments and suspend or modify some of them so that the remaining ones are those that fit one's

best principles and theoretical framework. Reflective equilibrium, for Rawls, “is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation.”²⁴

Margaret Urban Walker’s distinctive version of reflective equilibrium is situated in her “expressive-collaborative” model in which morality *consists* in human social practices and cannot be separated from other social practices.²⁵ For Walker, critical reflection on morality is also deeply social; thus instead of devising a theoretical model for reflection—which can go on inside one person’s head—one must look at how actual communities engage in the social act of reflection. The status of intuitions in more Rawlsian versions of reflective equilibrium worries Walker because, while intuitions serve as “data” for the construction of theory, “there are no principled procedures for disqualifying moral data short of the moral theories that the data are supposed to constrain.”²⁶ The intuitions are those of the moral theorist, who may exercise discretion “not only in fitting theory to data, but in *fitting data to theory*.”²⁷ This leaves no space for an examination of why we (the theorists) might have the intuitions that we have; it allows no critical challenges from other people who are differently situated. Because “those moral claims will be ‘intuitive’ that we have learned to make in common with others who have received a like moral training or inhabit ‘our’ moral world”²⁸—and because the intuitions may have ideological support—it is not surprising that they would strike “us” as compelling; however, their being compelling should not be taken as evidence that they are valuable. One must ask how one’s social position has affected one’s moral training and thus which intuitions one has developed. If the origin or support for one’s intuitions is problematic—if, for instance, the intuitions require overlooking how one’s moral practices function in relation to other social structures—one should become suspicious of them.

Thus the method of reflective equilibrium should, according to Walker, facilitate a particular kind of scrutiny of intuitions. The equilibrium must be “*equilibrium between people as well as within them*,”²⁹ indicating that participants in a socially produced moral order must have the opportunity to critically disequilibrate the shared order. The process is an actual process of “social negotiation in real time, where members of a community of roughly or largely shared moral belief try to refine understanding, extend consensus, and eliminate conflict among themselves.”³⁰ In this process, what moral communities are doing is inquiring into the value of moral practices, or in Walker’s words, “assessing the *habitability* of a particular form of moral-social life.”³¹

One cannot assess the value of a moral practice, however, under all conditions. If a moral practice appears to be something other than what it is then assessments of its value will be unreliable. The ideological shaping of an intuition is achieved by making a practice appear to be something other than what it is: for instance, a gendered division of labor might appear to be giving men and women different but equally valuable and complementary spheres of work and of control, and

this appearance can trigger a positive intuitive evaluation of the practice; however, this depends on hiding some of its features from view, such as the way it creates unequal exit options for women and gives more bargaining power in the relationship to men. If these features were visible then one's intuitive responses (especially women's intuitive responses) to the practice would more likely be negative.³² The key, then, is for the process of reaching reflective equilibrium to include a strategy for making moral practices "transparent"—making all of their features visible—and for assessing them once they are transparent; Walker calls this strategy "transparency testing." What makes a state of equilibrium properly reflective is this form of critical reflection that "presses toward transparency."³³ What is special about transparency testing as a strategy for the construction of authoritative values is that it is attentive to the operations of power that gave rise to the initial intuition. Intuitions are suspect because they are likely to have been ideologically shaped so as to mask the fact that they support "uninhabitable"³⁴ forms of life for certain people.

Transparency testing can lead people to lose confidence in a moral practice, and this is important because participants in a moral practice, or people who share a certain moral understanding, must have confidence in it in order for it to gain or to maintain what Walker calls "the specifically moral authority of morality."³⁵ However, confidence will tend to be misplaced when moral practices appear to be something other than what they are. A practice that is revealed through transparency testing to depend on force or coercion will fail, Walker argues, to continue to command confidence, and this loss of confidence strips the practice of its moral authority.³⁶

Walker's approach is Neurathian, so the scrutiny of any values takes place from the vantage point of other values; namely, those that one (at least for now) holds more confidently and that have already attained a state of reflective equilibrium. Hence the question to ask about a particular moral understanding that has been made transparent is something like: does it conflict with our other, more confidently held values? For instance, feminist ethicists situated within Western democracies presume some shared values "of specifically democratic, participatory, and egalitarian kinds, squarely founded on moral and political ideals of modern Western social thought"; within this context they can insist that the values associated with a particular moral practice be consistent with these already shared values, for revealing inconsistent values "magnifies embarrassing double-binds of modern morality."³⁷ Equilibrium is then disturbed, until members of the moral community renegotiate and reject the practice that one might describe as having failed the transparency test.³⁸ To subject a practice to transparency testing, then, is to at least consider rejecting it, for one commits to rejecting it if it creates an "embarrassing" contradiction with one's other values.

To summarize: For Walker the construction of morality—that is, the process through which moral claims gain or lose "the specifically moral authority of morality"—must include transparency testing in order to weed out ideologically

shaped intuitions. This is done by using reflective equilibrium to see which of one's intuitive judgments (once they have been made transparent) are and are not consistent with one's other, confidently held values. Every intuition must be scrutinized in reflective equilibrium, so every intuition must be subjected to *reasoned* reflection, for only a reasoning process can identify and eliminate inconsistencies. I am questioning this standard picture of construction as always requiring reasoning. The standard picture fails to recognize that reasoning is not the only (or even the primary) cognitive process through which humans as valuing creatures do what I take to be what construction is really about: the creation of values that one affirms from within one's own evaluative standpoint (or what Street calls one's "practical standpoint"), or an evaluative standpoint that is shared by members of one's moral community more generally.

Thus my response to Walker's prescription for transparency testing and for the kind of reflective equilibrium that it enables is a "Yes, but . . ." sort of response that comes from having an eye on value pluralism and its implications for meta-ethical constructivism; I mean here not just the fact that there can be plural and conflicting values, but also the fact that there are a plurality of cognitive *processes* through which valuing or evaluative judging can take place and, I would add, a plurality of cognitive processes through which a second-order affirmation or endorsement of one's judgments can take place. There are, I am claiming, a plurality of ways in which one might produce the "specifically moral authority of morality." We have morality because we *value*,³⁹ and we do much of this valuing in an affect-laden, intuitive way. Furthermore, some of our moral judgments—those regarding sacred values—*must* be made intuitively. My worry is that the process of construction—if it is conceived as a form of reflective equilibrium, including Walker's version—can degrade a sacred value, precisely because of the way that it requires reasoning about whether or not to eliminate each of one's evaluative judgments. I contend that sacred values should not be subjected to this method, and suggest that there are other ways in which they might be affirmed and imbued with authority. Sometimes it is best to allow a moral judgment's authority to be constituted by the fact that one's confidence in it is such that it *resists* scrutiny.

To see the problem with limiting the process of construction to the method of transparency testing (or other versions of reflective equilibrium) in the case of sacred values, consider the following example: if one were to ask, "Must I protect my child?" the answer is simply a confident "Yes—the alternative is unthinkable," rather than "Yes, since I am confident about this even after checking to see whether my sense that I must protect her withstands scrutiny from the point of view of my other values." The latter answer would manifest *one thought too many*. It is not that I am worried that my judgment ("I *must* protect my child") would be found to be mistaken if I were to scrutinize it from the point of view of my other deeply held values; rather, it should not be subjected to any such reasoned consideration, lest I think the unthinkable. Engaging in the kind of scrutiny that is usually taken to be an essential feature of constructivism implies that one thinks that each judgment

being tested could potentially be rejected—and that is precisely what *must not be thought* in the case of sacred values.

If sacred values are to be preserved, they must be permitted to take a different route to attaining the “specifically moral authority of morality” than more negotiable moral values take. My confidence in the authority of the moral requirements associated with a sacred value does *not* depend on my finding, through a reasoning process, that they are consistent with my other values. Instead, my confidence in them must be, like the initial judgment, automatic. Reasoning about them would reveal a *lack* of the automatic confidence that they call for. And, even if reasoning did not destroy these values, the discovery of a conflict with my other confidently held values would not indicate that I should rid myself of either value: if values are plural in the way that I take them to be, then I might be fully confident in *both* of two conflicting values.⁴⁰

My confidence that I must, for example, protect my child must be *decisive* in such a way that it cuts off further consideration. Harry Frankfurt’s account of seeking second (and higher)-order evaluations—which one does because of what he calls the *reflexivity* of human consciousness—illuminates this phenomenon of decisiveness. For Frankfurt, someone who has only first-order desires and does not choose from among them which to make her/his will, but rather acts on whichever is strongest at any given moment, is a “wanton.” Most people are not wantons, because they not only have first-order desires and second-order desires, which are desires about which desires to have, but they also have what Frankfurt (1971) calls second-order volitions, which have to do with what one wants to be one’s *will*.⁴¹ It seems that it would be one’s highest-order volitions that matter. But Frankfurt identifies a problem with the simple idea that there is a hierarchy of desires and volitions: there can be an infinite regress of wantonness, because one can always take one’s reflexivity up to a higher and higher level.⁴² One might have second-order volitions about one’s first-order desires, but if these second-order volitions conflict, then one needs third-order volitions to choose between one’s second-order volitions, and so on. What puts an end to the infinite regress is decisiveness, or more precisely, one’s decisive identification with one of one’s volitions: “such an identification resounds through an unlimited sequence of possible further reconsiderations of [one’s] decision.”⁴³ Decisive identification takes place when one has confidence enough to make a commitment, and

a commitment is decisive if and only if it is made without reservation, and making a commitment without reservation means that the person who makes it does so in the belief that no further accurate inquiry would require him to change his mind. It is therefore pointless to pursue the inquiry any further.⁴⁴

At the moment that one makes a decisive commitment, the regress stops. The process of construction, as it is standardly construed, can be thought of as a process for eliminating conflict among one’s first-order evaluative judgments and arriving at a reasoned, higher-order endorsement of only some of those judgments. In

such a process, there can always be reconsiderations; the Neurathian procedure continues without end. What Frankfurt describes as a decisive commitment is an alternative to this process of reasoning one's way to the resolution of conflicts, and it is an alternative that still allows one to imbue an initial evaluative judgment with the kind of authority that the products of construction have.

Elsewhere Frankfurt (1982) has argued that one who loves will experience the value of the beloved as compelling, or as generating the necessity—the *requirement*—of the actions that are partly constitutive of loving, such as actions that protect, preserve, care for, nurture, or repair the beloved; Frankfurt calls these *volitional necessities*. As Frankfurt puts it, “an encounter with necessity of this sort characteristically affects a person . . . by somehow making it apparent to him [*sic*] that every apparent alternative to that course is unthinkable.”⁴⁵ The person who affirms a volitional necessity “accedes to it because he is *unwilling* to oppose it and because, furthermore, his unwillingness is *itself* something which he is unwilling to alter.”⁴⁶ Though Frankfurt does not point this out, I believe that the affirmation of a volitional necessity is a clear case of a decisive commitment that “resounds” through any potential further considerations. It decisively stops whatever regress there might otherwise be of seeking higher and higher-order affirmations of an evaluative judgment. The moment one grasps a volitional necessity—the moment one apprehends or experiences one option as nonnegotiablely required and alternatives as unthinkable—one rules out any further reflection that could end in a rejection of what one takes to be a sacred value.

What I take to be most significant about certain instances of decisiveness is that they are reached through automatic processes and shield one from thinking the unthinkable.⁴⁷ In contrast, the method of reflective equilibrium treats the apprehension of volitional necessities—that is, of nonnegotiable requirements whose alternatives must be automatically ruled out as unthinkable—as no different from any other intuitive evaluative judgments: they all must be subjected to reasoned reflection. Frankfurt thus highlights (though this is not his aim) a route to confidence that is quite different from reflective equilibrium or from Walker's transparency testing, but that I would still consider to be a process of construction. The process of construction that starts with an intuitive judgment about a sacred value is itself an intuitive and automatic process, and yet it is decisive. The initial intuitive judgment, “I *must* A” or “A is required,” is accompanied by a decisive commitment that consists of the intuitive realization that “the requirement to A is non-negotiable, and the alternative to A is unthinkable.”

IV

I have found—borrowing from Frankfurt—a route to the endorsement of sacred values, but I am still left with a problem: it is vital to have a way to *reject* values that have been sacralized, because sometimes people sacralize values badly. Walker

is right to worry about ideologically shaped intuitions, and I believe that sacred values can be subject to this ideological shaping just like other values can, and thus they can trigger people's confidence by appearing to be something other than they are; that is, by lacking transparency. Were I to engage in critical reconsideration of a sacred value, I might actually decide that it was bad to have sacralized the value, but such reconsideration will count as a transgression of the value if it turns out that it "really" is sacred—I will have had one thought too many. I am in a double bind: if the value "really" is sacred, I must not reconsider it, and must instead have automatic, decisive, confidence in it, but if it is "wrongly" sacralized, then it seems that the only way to discover this is to engage in reasoned reflection and reconsider it. Thus the only route that I might take toward rejecting a value that I have sacralized is analogous to a "trial by water" in which I must risk drowning the innocent. According to this (supposed) method of identifying witches, to confirm the innocence of an alleged witch one must drown her, for anyone who floats is thereby shown to be a witch. Unfortunately, in a trial by water, in order to be free of the (purported) threat of witches, one ends up killing off the innocent; similarly, to be free of the risk of "wrongly" sacralized values, one ends up transgressing or degrading all of one's appropriately sacralized values.⁴⁸ If one refuses to take the risk of transgressing a sacred value, then one risks instead that one is unwittingly sacralizing some values that one should not sacralize. Thus the alternative to reflective equilibrium that I have proposed in the case of values whose rejection is unthinkable—namely, imbuing an intuitive judgment with authority through one's decisive confidence in it—could be said to place too much trust in one's intuitions. As Simon Blackburn has pointed out, while consequentialism may direct one to think "one thought too many," it is also possible to have "one thought too few."⁴⁸ My account (so far) seems to encourage this.

I do see one way out of this double bind, and it has to do with the fact that—as Walker emphasizes—the process of reflective equilibrium is a social process. I think that Walker's transparency testing can, without any problematic distortion, be reconceived along the lines of Haidt's social intuitionist model, which gives a much less central role to reasoning and a more central role to automatic, intuitive processes. I suggest that when the process of the construction of moral values is reconceived in this way, it reveals a route to the rejection of judgments about values that have been sacralized by some, but not other, members of a moral community. When other people do not share my intuitive judgments about what is and is not unthinkable, *they* can take themselves to be free to engage in reasoned reflection about whether to reject a value that *I* treat as sacred; *they* can do the thinking that I have judged it to be impermissible for me to do. In fitting transparency testing to Haidt's social intuitionist model, I am insisting that it is not necessarily a process of reasoned reconsideration of one's own values. Instead, if one is a member of a moral community in which different people sacralize different values, then interaction with different people might take the place of intentionally testing (and thereby transgressing) one's own sacred values in a trial by water. *Other people* can

throw my sacred value in the water for the purpose of demonstrating to me that it will float (and that I should thus reject it).

Recall that according to Haidt, once an intuitive judgment has been made by an individual, that individual's own reasoning—which is “motivated” to support her/his intuitive judgment—is unlikely to result in a change in the judgment. However, if the individual interacts with other people who do not share that judgment, the other people may provide reasons (which they developed *post hoc* to support their own prior intuitive judgments) that oppose the individual's judgment, or they may, instead of providing reasons, simply reframe an issue—and this may trigger new intuitions in the individual. The social interactions that result in people's changing their judgments do not have to involve reasoning at all; simply the fact that other people whom one trusts make a particular judgment exerts an influence on one's own judgment.⁵⁰ It is largely through having a particular *affective* experience in one's interactions with others—rather than through reasoning—that one comes to change one's moral understandings.

Thus while Walker is right that the moral understandings that support pernicious power structures are largely intuition based, it is also largely affect-laden intuitions that can ultimately *undermine* these structures. If this is correct, then transparency testing is best characterized as taking place by people (for instance, in feminist communities) prompting new intuitions in each other—namely, the intuitions that arise when one begins to see a moral practice transparently. This is a good depiction, for instance, of what typically took place in the feminist “CR” (consciousness raising) groups that were a vital part of the second wave of the women's liberation movement. The confidence that one has after transparency testing may be, after all, a confidence in one's (new) intuitions, rather than a confidence that rests on the reasoned judgment that a particular moral understanding is consistent with one's other values. This social intuitionist version of transparency testing can still be thought of as a Neurathian process, but a social one, and one that relies heavily on affective interactions and automatic processes. It is not I, but rather we, who are on board the Neurathian ship, and this “we” may be diverse. I might not be able to step off the rotten plank of the ship that I am standing confidently on—it might even be unthinkable for me to do so—and so I might be unable to see, from the viewpoint of other planks, that it needs repair or replacement. But other people stand on other planks, are confident of other things, and can trigger changes in me and my values.

Though people can in this way provide a check on each other's intuitions, the opportunity for this check may disappear if one expresses one's values only with other people who have had similar kinds of experiences and moral training; the plurality and diversity of values in a moral community is crucial. What Walker's version of reflective equilibrium emphasizes is that it matters very much whom one talks to and where they are situated in power structures. It matters whether I talk with others who have a stake in preventing a particular moral practice from becoming transparent (generally those who are beneficiaries of a practice), or

whether I talk to those who are subordinated through the practice. If I have treated as sacred some value that creates an “uninhabitable” kind of life for myself or for others, there is a chance that encountering others’ intuitions that are contrary to my own will prompt me to see this.

I am not claiming that I have somehow eliminated the risk posed by the sacralizing of some values. It remains the case that I may treat a value as sacred and insulate it from scrutiny, refusing to even consider rejecting it, and resisting all attempts by others to try to provoke me to reconsider. So the problem remains: some moral judgments must only be made intuitively, for to reason about them is to destroy them; however, intuitions can be dangerous, so critical scrutiny is required. If I’m not willing to give up my sacred values—which I’m not—then I must accept instead that preserving them simply makes moral life risky. In the next section I will try to make sense of why, even given the risks, I remain decisively confident that certain things are unthinkable, and will say what I think is good about having this kind of decisive confidence about some values.

V

I must be understood to be speaking from aboard a Neurathian ship—engaging in an evaluative activity without a foundation—when I express my response to this normative question: is it good to sacralize some values and protect them from one’s own critical reasoning (even if they are not protected from other people’s criticism)? I’ll give two conflicting answers: Yes—human attachment requires that one sacralize some values and then protect them, and this attachment is both inevitable (in the sense that it is part of human nature) and desirable (in fact, I take it to be the best thing in life); relationships of love or of care for particular others require sacralizing the value of one’s beloved or of the one for whom one cares. But also: No, it is not all good, in that there are costs. I will explain the “yes” answer first.

Very simply, I believe that not taking some things to be unthinkable would entail giving up an important value, the source of which is love or attachment. This is not the only value, or the only value that can be moralized, but it is a central one; value is constructed by valuing creatures, and so it is important that attachment gives rise to potent experiences of valuing. These experiences may involve intensely protective and nurturing responses to those whom one loves or for whom one cares, and great distress when those whom one values are threatened or suffering. The automatic, intuitive judgments that come with human attachment have parallels in the automatic processes of other mammals, who lack the neural systems for the reasoning in which humans can additionally engage. Patricia Churchland (2011) has detailed some of these processes, including how such things as the release of oxytocin enable trust and attachment; this in turn

underlies the emotional responses that occur when, for instance, a loved one is threatened.⁵¹ Attachment has its roots in self-preservation, but the impulses involved in self-preservation get extended to offspring and then out to increasingly large circles. All mammals have this response to some degree, but some species more than others, and humans, as a particularly social species, have it to a particularly high degree. I assume that part of what explains my confidence in some of my intuitive judgments is simply that I am attached—strongly enough that certain things become unthinkable.⁵²

Of course, not everything that is part of human nature is good or yields authoritative reasons. I am not claiming that it is *because* attachment and its accompanying evaluative judgments are natural that they are therefore good. Were I to do so, I would commit the fallacy of deriving an “ought” from an “is,” or a prescription from a description. As expressivists have taken pains to emphasize, when I express an evaluative attitude (as I have just done), I am doing something different from describing, though from a third-person perspective, I could be described as having a particular evaluative attitude, and I and my evaluative attitude are parts of the natural world. Because I take the expression of my evaluative judgment to be part of the social, Neurathian process, I am relying on that process to sort out which of my (“natural”) evaluative judgments to keep and which to try to overcome. In this way, one *constructs* rather than *derives* an “ought,” and one does so not from a description of what is, but rather from and through one’s evaluative activities.

There may be aspects of human nature and associated evaluative judgments that should be identified as causing problems—as disvalues—and that should be stifled, while other aspects are to be encouraged and enlarged. Martha Nussbaum, for instance, argues that people should try to overcome “projective disgust”; namely, “disgust for a group of other humans who are segmented from the dominant group and classified as lower because of being (allegedly) more animal.”⁵³ Meanwhile, she sees love (defined, for her, as “intense attachments to things outside the control of our will”) as so valuable that she is led to claim that “all of the core emotions that sustain a decent society have their roots in, or are forms of, love . . . Love . . . is what gives respect for humanity its life, making it more than a shell.”⁵⁴ Nussbaum thus approaches the study of the evolutionary origins of those aspects of human nature relevant for morality as revealing both “the resources and the problems that human nature, insofar as we can know it, makes available.”⁵⁵

What, then, about the kind of love that comes with evaluative judgments that certain things are unthinkable? I would not want to live in a world in which this form of valuing—and a decisive commitment to it—were absent. Frankfurt makes a related point about love (which, for him, comes with volitional necessities):

What we love is necessarily important to us, just because we love it. There is also a rather different point to be made here. Loving itself is important to us. Quite apart from our particular interest in the various things that we love, we have a more generic and an even more fundamental interest in loving as such . . .

Why is loving so important to us? Why is a life in which a person loves something, regardless of what it is, better for him [*sic*] . . . than a life in which there is nothing that he [*sic*] loves? Part of the explanation has to do with the importance to us of having final ends.⁵⁶

My confidence in my intuitive judgments about nonnegotiable moral requirements depends on my valuing *having* the kinds of relationships that give rise to these affect-laden, intuitive judgments. I imagine a world without sacralized values (and their accompanying intuitive judgments) as a dystopia in which no one could be adequately loved in her/his particularity; instead, we would all be like the wife of the rescuer who, by having one thought too many, reveals that even if he finds it to be *wrong* to not prioritize his drowning wife over a drowning stranger, he does not find it to be *unthinkable*.

While a world in which one cannot have this strong kind of particularized love is already dystopian, it is more than just this kind of love that would be absent in a world in which nothing were unthinkable. Consider that most people, if faced with just a drowning stranger (and not, as Williams's rescuer was faced with, a stranger and one's wife), would find it unthinkable not to rescue that stranger. In the dystopia in which nothing is unthinkable, people would also lack the kinds of intuitive judgments that, in the real world, most people have in certain situations in which proximate strangers are in great danger.

However, one cannot rely on one's intuitive judgments about what is unthinkable to identify all nonnegotiable moral requirements. For instance, while a proximate, particular stranger whose life is in danger will tend to trigger an intuitive judgment that it would be unthinkable not to respond, distant, anonymous strangers whose lives are in danger will tend not to trigger a similar judgment.⁵⁷ One can try to purposefully arrange situations in such a way that they *do* trigger automatic judgments; for instance, charitable organizations do this in their solicitations by telling the story of just one particular individual who is suffering from a disaster, rather than reporting a statistic about how many people are suffering. Because of the malleability and unreliability of our judgments about what is and is not unthinkable, sacred values, and one's decisive confidence in them, cannot be the *only* basis for the authority of moral requirements. Some other kind of moral value may do a better job at motivating a response to those whose needs do not tend to trigger an automatic judgment of unthinkableability.

There may be a temptation to recognize *only* these other, typically impartial, kinds of values—such as the value of fairness or of maximizing some good—as *moral* values, and to count values involving love or other kinds of care as non-moral. I reject this narrow account of what morality includes, and instead employ the functionalist and pluralist account of morality mentioned at the beginning of this article, according to which morality is a set of values (and associated requirements, norms, practices, traits, structures, and so on) that contribute to some form of social connection or cooperation and that have come, through an ongoing social process, to be understood as carrying a special kind of authority. Given such

a functionalist and pluralist account of morality, one can recognize both impartial values and the values involved in caring for particular others as moral.⁵⁸ This has long been accepted in (feminist) care ethics, and I follow this tradition here. A narrow account of morality may not acknowledge that a plurality of values that connect people in ways that facilitate cooperation and balance self-regarding and other-regarding concerns can be moralized, that is, can be valued and imbued through a process of social construction with a special authority. Here Frankfurt and I part ways, because he classifies the requirements of love and care as non-moral requirements.⁵⁸ This is because he does have a narrow conception of morality—for instance, he sees morality as impersonal⁶⁰—and he wants room for other kinds of normativity than morality, seeing these other kinds of normativity as sometimes overriding (narrowly conceived) morality.⁶¹

I do want to emphasize, however, that intuitively grasped moral requirements that have their roots in attachment cannot be *all* of morality—and this grounds my “no” answer to the question of whether it is good to sacralize some values, treat their sacrifice as unthinkable, and exempt them from having to withstand a certain kind of scrutiny. One worry that I have is that sacralizing some values can crowd out other important values. Thus I reject claims by those such as Churchland, who views attachment and its associated automatic responses as the *only* basis for morality. She claims: “Attachment, underwritten by the painfulness of separation and the pleasure of company, and managed by intricate neural circuitry and neurochemicals, is the neural platform for morality.”⁶² I would modify this: moral values and their sources are plural, so the neural platform that supports attachment is *one* of the (crucial) neural platforms for morality, and perhaps it is what best explains the sacralizing of values and the affective responses that lead one to perceive some things as unthinkable. But there are other moral values, and requirements arising from them may be in tension with sacralized values.⁶³

A concern with oppression makes some of these tensions particularly salient. Consider Walker’s worries, for instance, about the way that oppressive structures conceal features of some moral practices and moral understandings, and garner people’s confidence precisely by appearing to be something other than what they are. Relationships of love or other kinds of care—which are typically highly gendered—could be oppressive while at the same time triggering the intuitive grasping of nonnegotiable moral requirements, where these intuitions are shaped by the same oppressive structures that sustain the relationship. Attachments are not always good or healthy. A battered woman may love and care for the batterer and intuitively judge that she is nonnegotiablely required to remain loyal to him and to defend him against others. If moral requirements such as these are exempted from the woman’s own critical scrutiny because for her to subject them to such scrutiny would constitute having “one thought too many,” then their oppressive elements may remain undetected, especially because, given the typical dynamics of abuse, she may be isolated from other people who would perceive her as having “wrongly” sacralized the value of her relationship to the batterer. Here the value

of freedom from oppression or domination can be seen to conflict with a sacred value.

Other values, such as justice or fairness, can also compete with the value of automatically taking some nonnegotiable moral requirements to be authoritative. Justice plays an important role in moral requirements about how to treat others whom one regards as “distant” or as members of an out-group, or any others toward whom one does not *automatically* judge oneself to have nonnegotiable moral requirements.⁶⁴ On the one hand, intuitions about what is just or fair need not be all that lead one to recognize requirements toward distant others, and care-based values, including those that have been sacralized, can be extended toward distant others depending on how one directs one’s attention to them; Nussbaum’s suggestion that one enlarge one’s “circle of concern” and extend some form of love further out, enriching one’s grasp of what it is to respect humanity, assumes that love can be the basis of one’s recognition of nonnegotiable moral requirements toward distant others.⁶⁴ On the other hand, fairness also seems to play an indispensable role in how moral requirements toward distant others are conceived, and judgments about the fair treatment of others are at risk of being preempted by judgments involving sacred values, in any case when actions required by fairness and actions required for securing sacred values would conflict. This, too, is a serious risk. Thus my pointing to the importance of the automatic responses that are related to attachment should not be taken to support the sacrifice of potentially conflicting values that *are* best achieved through reasoning, such as the value of fair and impartial treatment of others.

I am suggesting that *neither* a prescription to simply act on all of one’s moral intuitions nor a prescription to always engage in controlled reasoning is adequate, for neither will lead one to affirm enough of the great plurality of values that seem to be vital for a rich human morality. One might think, then, that the key is to know when to rely on intuition and when to exercise control and engage in reasoning.⁶⁶ However, I do not believe that, in all situations, there is necessarily a good moral choice available, and that one could reliably arrive at this good choice if only one knew how to distinguish situations that call for intuition from situations that call for reasoning. The problem is that in some situations there are plural, and often conflicting, values to be affirmed through different processes, and in some of these cases, no one of the conflicting values can substitute for the other value without unique—and perhaps tragic—loss. These can be situations of unavoidable moral failure.⁶⁷ Indeed, the sort of moral value pluralism that I have been discussing implies that moral life can be dilemmatic, through and through.

ACKNOWLEDGMENTS

This article contains extensive excerpts from Lisa Tessman, *Moral Failure: On the Impossible Demands of Morality*, Oxford University Press (2015). The ideas and arguments in this article are developed more thoroughly in the book, and are also connected in the book to the claim that—given the intuitive, automatic, cognitive process that produces many of our moral judgments, particularly judgments about the sort of nonnegotiable moral requirements discussed here—there are some moral requirements that contravene the principle that “ought implies can,” and thus that can make moral failure unavoidable; I encourage readers who are interested in this line of thought to see how it is developed in the book.

NOTES

1. Human sociality is complex and takes place at many levels, in part because we are socially connected (or not) so differently to different people; for instance, to particular, intimate others including our children, to other kin, to other members of a variety of groups to which we may belong, and to distant strangers. Many of our values and ways of valuing are evolved mechanisms, and the fact that there is multilevel selection, and that both biological and cultural evolution play a role in shaping our values, in part account for the plurality of moral values. My account here is largely consistent with (and indebted to) David Wong’s: “If individual and group forms of natural selection have shaped human nature, then an intuitively plausible result is not only pluralism of motivation but profound ambivalence in human nature. The force of selection on the individual can be expected to have produced a strong dose of egoistic motivation, a significant capacity to take into account the welfare of others and to contribute to that welfare at varying levels of cost to the self, and various extremely familiar and frequently occurring motivations that seem to fall in between pure forms of egoism and altruism. Moreover, a mixture of such motivations is probably the norm for human beings . . . Rather than saying that an effective morality should always constrain self-concern and reinforce other-concern, it should often attempt to accomplish a productive balance or reconciliation between those types of concern . . . The different mechanisms that result in the various forms of altruism—kin, group, sexual and cultural selection, and reciprocal altruism—help to explain the plurality of basic value” (2006, 57–59). I also make use of Margaret Urban Walker’s idea of the “specifically moral authority of morality” (2003, 109), which I will discuss further.
2. I take the position of moral value pluralism to consist in the following three claims: (1) there are plural sources of moral value, which give rise to a plurality of kinds of moral values (and correspondingly plural kinds of moral requirements); (2) moral values that differ in kind may be incommensurable with each other, which suggests that judgments of the relative worth of values of two different kinds cannot be made by measuring the quantity of some common property found in each; and (3) values of different kinds are nonfungible, so when a value of one kind is chosen over a value of a different kind, the chosen value does not substitute for the loss of the rejected value. Wong writes that “‘Moral value pluralism’ is the doctrine that there exists a plurality of basic moral values, where such values are not derivable from or reducible to other moral values” (2006, 6). For other discussions of moral pluralism and its implications, see Lemmon 1962; Nagel 1979; Raz 1986, chapter 13, and 2003; Stocker 1990; Anderson 1993.
3. Or what Harry Frankfurt calls the *commands of love*; see Frankfurt 2004.
4. Street 2010, 367. A survey of different versions of constructivism can be found in the introduction and in some of the essays in Lenman and Shemmer 2012.

5. Street 2008, 208; italics in the original.
6. *Ibid.*, 223; italics in the original.
7. Otto Neurath's image of sailors was used by Quine to metaphorically describe the nonfoundational approach to critiquing and revising one's "conceptual scheme": "We must not leap to the fatalistic conclusion that we are stuck with the conceptual scheme that we grew up in. We can change it bit by bit, plank by plank, though meanwhile there is nothing to carry us along but the evolving conceptual scheme itself. The philosopher's task was well compared by Neurath to that of a mariner who must rebuild his ship on the open sea. We can improve our conceptual scheme, our philosophy, bit by bit, while continuing to depend on it for support; but we cannot detach ourselves from it and compare it objectively with an unconceptualized reality" (Quine 1963). Walker argues explicitly for a parallel between Quine's naturalized epistemology and a naturalized moral epistemology, though she is also careful to point out aspects of Quine's epistemology that moral epistemology should reject (see Walker 2003, chapter 11). Walker also comments: "We always stand on some moral values as we consider the authority of others" (2003, 110). Yonatan Shemmer has a category of constructivist that he dubs "Neurathian constructivist" (2012, 162), mentioning Lenman as an example (see Lenman 2012, 217; see also Lenman 2007, 76, and Lenman 2010, 181). See also Gibbard 1990, 106.
8. Kahneman 2011, 20–21.
9. Jonathan Haidt (2001, 818) provides the following table to contrast the features of the two systems (see Haidt for citations of the research that this table summarizes):

The intuitive system	The reasoning system
Fast and effortless	Slow and effortful
Process is unintentional and runs automatically	Process is intentional and controllable
Process is inaccessible; only results enter awareness	Process is consciously accessible and viewable
Does not demand attentional resources	Demands attentional resources, which are limited
Parallel distributed processing	Serial processing
Pattern matching; thought is metaphorical, holistic	Symbol manipulation; thought is truth preserving, analytical
Common to all mammals	Unique to humans over age 2 and perhaps some language-trained apes
Context dependent	Context independent
Platform dependent (depends on the brain and body that houses it)	Platform independent (the process can be transported to any rule-following organism or machine)

10. Conflict between the two systems arises, for instance, in the Stroop color-naming task (Stroop 1935), in which one must name the color of the ink in which a word is written, where the word is the name of a color, but the color named by the word and the ink color are mismatched (e.g., "red" written in purple ink, "blue" written in green ink, etc.). One's automatic response is to say the word that is written (assuming fluency and reading ability), so in the case where word and ink are mismatched, the verdicts of System 1 and System 2 come into conflict; one's automatic response must be effortfully overridden so that one can name the ink color, and this engagement of System 2—the exercise of cognitive control—takes attentional resources and slows response time (in comparison, when viewing an ink square instead of a color word written in ink, one can name the ink color without conflict, and thus more quickly). Because System 2 must allocate the limited resources that it requires to operate, when it is engaged with one task, it will perform another task less well or more slowly.
11. Cushman, Young, and Greene 2010, 48.

12. Haidt 2001; 2012.
13. Additional evidence for the claim that conscious reasoning usually occurs after rather than before a moral judgment is made can be found in the phenomenon of what Haidt calls “moral dumbfounding” (Haidt 2001; 2012). Moral dumbfounding takes place when a subject makes a judgment (for example, that it is wrong to eat one’s dead pet dog) and then is unable to come up with a reason to support the judgment, but nevertheless remains unshaken in her/his commitment to the judgment. Had conscious reasoning been what led the subject to the judgment in the first place, such moral dumbfounding would not take place—that is, the reason would still be readily available to the subject. There is much additional evidence to support the hypothesis that reasoning takes place *post hoc* to rationalize intuitive moral judgments; see Haidt 2001 and Haidt and Kesebir 2010 for this evidence.
14. See also Ditto et al. 2009.
15. Haidt 2001, 819. Reasoning that is *not* social or interpersonal and that changes one’s own intuitive judgments is rare, but does occur. Haidt posits two ways in which one might reason one’s way to a new moral judgment. The first way directly links reasoning with moral judgment: “people may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition . . . However, such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high” (Haidt 2001, 819). The second way indirectly links reasoning to moral judgment; the reasoning—for instance, reasoning in which one imagines oneself in someone else’s shoes—triggers “a new intuition that contradicts the initial intuitive judgment” (Haidt 2001, 819) and then this conflict of intuitions must be resolved to produce the final moral judgment.
16. Tetlock et al. 2000, 853.
17. *Ibid.*, 858.
18. *Ibid.*, 856.
19. *Ibid.*, 860; see also Fiske and Tetlock 1997, Tetlock 2003, and McGraw et al. 2003.
20. Raz 1986.
21. Tetlock et al. 2000, 854.
22. Williams 1981, 18. Tamar Gendler (in a presentation at the New York University Center for Bioethics’ conference on “The Moral Brain,” March 2012) has also connected Williams’s case of the rescuer who has “one thought too many” with Philip Tetlock’s work on “sacred values.” Thanks to Gendler for calling my attention to Tetlock’s work.
23. Elsewhere (2015) I have argued that the rescuer was in a genuine moral dilemma, such that all of the available options would constitute moral failures: it is wrong not to automatically give preference to rescuing one’s beloved, but it is also wrong not to rescue the stranger. I argue that neither of these two conflicting moral requirements are governed by the principle that “ought implies can.”
24. Rawls 1971, 20. The equilibrium is considered to be “narrow” when convictions and principles have been found that are consistent with a set of particular judgments while requiring few revisions to this set, and “wide” “when someone has carefully considered alternative conceptions of justice and the force of various arguments for them” (Rawls 2001, 31). One can also speak of an equilibrium between people rather than within one individual, and this can potentially be a “full” reflective equilibrium (Rawls 2001, 31). See also Daniels 1979.
25. Walker 1998, 14.
26. *Ibid.*, 64.
27. *Ibid.*, 64.
28. *Ibid.*, 67.
29. *Ibid.*, 65; italics in the original.
30. *Ibid.*, 64.
31. *Ibid.*, 214.
32. *Ibid.*, 65.
33. *Ibid.*, 11. Walker cites Bernard Williams’s call for transparency in the ethical aspects of a society: “the workings of its ethical institutions should not depend on members of the community misunderstanding how they work” (Williams 1985, 101).

34. *Ibid.*, 214.
35. Walker 2003, 108.
36. *Ibid.*, 109.
37. Walker 1998, 73.
38. One complication is that transparency is not easily attained. The development of feminist consciousness—in which the deceit and coercion that have been propping up certain practices become visible—may lead to a *loss* of confidence in what one was previously confident about, without any accompanying clarity about what one can become or remain confident in. Sandra Bartky has described a similar phenomenon with her wonderful phrase, “double ontological shock”: “[F]eminists suffer what might be called a ‘*double ontological shock*’: first, the realization that what is really happening is quite different from what appears to be happening, and second, the frequent inability to tell what is really happening at all” (Bartky 1990, 18). Transparency testing might involve a “double *normative shock*” that leaves one without confidence about what to value. This is in part an epistemological difficulty—it is hard to know what one *would* have confidently valued, absent the deceit and coercion that were instrumental in shaping one’s actual confidence. For instance, a woman may not know whether she would have chosen a particular heterosexual relationship absent the coercive nature of the institution of “compulsory heterosexuality,” or whether she would have chosen to have children absent the ideology of motherhood. The difficulty is also in part due to the inseparability of pernicious and benign social constructions in any nonideal world: what is good and worthy of allegiance is all mixed up with what is not, so a single practice might command confidence in some ways and not in others. I take this complication to simply give rise to a sort of ambivalence that feminists, and all kinds of social critics, learn to live with.
39. Or as James Lenman puts it, “we are creatures who care about stuff” (2009, 1). Patricia Churchland asks questions that lead her to investigate values at the very most basic level: “Where do values come from? How did brains come to care about others? ... how is it that brains care about *anything*? To put it more tentatively, how can neurons care? What does it mean for a system of neurons to care about or to value something?” (2011, 12–13).
40. The implications of this claim are further developed in Tessman (2015).
41. Frankfurt 1971, 10–11.
42. Frankfurt 1988a, 165.
43. *Ibid.*, 168.
44. *Ibid.*, 168–69.
45. Frankfurt 1982, 263.
46. *Ibid.*, 264.
47. Whereas Frankfurt emphasizes decisiveness for the sake of showing that an otherwise infinite regress has a nonarbitrary stopping point.
48. It is a myth that “trial by water” was widely used on people who were accused of witchcraft; it was actually used very rarely (and there is some question of whether it was used at all).
49. See Blackburn 1998, 43–47.
50. Haidt 2001, 819.
51. Churchland 2011. She notes: “In using the word *attachment*, I am adopting the terminology of neuroendocrinology, where *attachment* refers to the dispositions to extend care to others, to want to be with them, and to be distressed by separation,” remarking that her clarification of this is meant to distinguish her use of the term from that of some clinical psychologists (2011, 16 and 208).
52. Recognizing the evolutionary origins of evaluative judgments, as I am doing here, does *not* serve a “dubunking” purpose for constructivist accounts of morality, as it would for realist accounts. See Levy 2006; Street 2006; and Kahane 2011.
53. Nussbaum 2013, 184.
54. *Ibid.*, 15.
55. *Ibid.*, 137.

56. Frankfurt 2004, 51–52.
57. This point is central to Peter Singer 1972, which has spawned a large literature.
58. See Wong 2006. Haidt also subscribes to a functionalist, and pluralist, account of morality: “Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate self-interest and make cooperative societies possible” (Haidt 2012, 270).
59. See Frankfurt 1988b, 182; 2004, 5–9; 2006, 46–48.
60. Frankfurt 1982, 268.
61. Frankfurt 2004, 9. Frankfurt further explains the relationship between morality and love by saying that we love our moral ideals, and that is why we become angry toward anyone who transgresses them (Frankfurt 2006, 46–48).
62. Churchland 2011, 16.
63. Haidt and colleagues have undertaken a research program to document and classify the plurality of moral intuitions that people (from many different countries and cultures) experience, calling their theory “Moral Foundations Theory.” They classify moral intuitions into six foundations: “care/harm,” “fairness/cheating,” “liberty/ oppression,” “loyalty/betrayal,” “authority/subversion,” “sanctity/degradation.” (In an earlier stage [Haidt and Joseph 2004], Moral Foundations Theory posited five rather than six foundations, leaving out “liberty/oppression.” See Graham et al. 2011, and see also Haidt 2012, which reviews earlier work on Moral Foundations Theory and discusses the addition of the sixth foundation.) While I would conjecture that these are not exhaustive and that there are other ways that one might carve up moral values (and I would not refer to them as “foundations”), this theory does give one a sense of the plurality and the range of values that have been moralized.
64. It is for this reason that my critique of the method of reflective equilibrium within metaethical constructivism is not a critique of Rawls’s “restricted constructivism” (Street’s [2008] term, to contrast with metaethical or thoroughgoing constructivism). Rawls limits the scope of construction to matters concerning justice, and the endorsement of sacred values would be inappropriate within this limited scope.
65. Nussbaum 2013.
66. This is essentially what Joshua Greene (2013) suggests.
67. See Tessman (2015) for the full development of this claim.

REFERENCES

- Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Bartky, Sandra Lee. 1990. *Femininity and Domination*. New York: Routledge.
- Blackburn, Simon. 1998. *Ruling Passions*. Oxford: Clarendon Press.
- Churchland, Patricia. 2011. *Braintrust: What Neuroscience Tells Us about Morality*. Princeton, NJ: Princeton University Press.
- Cushman, Fiery, Liane Young, and Joshua Greene. 2010. “Multi-Systems Moral Psychology.” In *The Moral Psychology Handbook*. Edited by John Doris. Oxford: Oxford University Press. 47–71.
- Daniels, Norman. 1979. “Wide Reflective Equilibrium and Theory Acceptance in Ethics.” *Journal of Philosophy* 76 (5): 256–82.
- Ditto, Peter, David Pizarro, and David Tannenbaum. 2009. “Motivated Moral Reasoning.” In *The Psychology of Learning and Motivation*, Vol. 50, Moral Judgment and Decision Making. Edited by B. H. Ross, D. M. Bartels, C. W. Bausman, L. J. Skitka, and D. L. Medin. San Diego, CA: Academic Press. 307–38.
- Fiske, Alan Page, and Philip Tetlock. 1997. “Taboo Trade-Offs: Reactions to Transactions that Transgress the Spheres of Justice.” *Political Psychology* 18 (2): 255–97.
- Frankfurt, Harry. 1971. “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68 (1): 5–20.

- Frankfurt, Harry. 1982. "The Importance of What We Care About." *Synthese* 53 (2): 257–72.
- Frankfurt, Harry. 1988a. "Identification and Wholeheartedness." In *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1988b. "Rationality and the Unthinkable." In *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 2004. *The Reasons of Love*. Princeton, NJ: Princeton University Press.
- Frankfurt, Harry. 2006. *Taking Ourselves Seriously and Getting It Right*. Edited by Debra Satz, with comments by Christine Korsgaard, Michael Bratman, and Meir Dan-Cohen. Stanford: Stanford University Press.
- Gibbard Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- Graham, Jesse, Brian Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter Ditto. 2011. "Mapping the Moral Domain." *Journal of Personality and Social Psychology* 101 (2): 366–85.
- Greene, Joshua. 2013. *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. New York: Penguin Press.
- Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814–34.
- Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage Books.
- Haidt, Jonathan, and Craig Joseph. 2004. "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues." *Daedalus* 133 (4): 55–66.
- Haidt, Jonathan, and Selin Kesebir. 2010. "Morality." In *Handbook of Social Psychology*, 5th ed. Edited by S. Fiske, D. Gilbert, and G. Lindzey. Hoboken, NJ: Wiley. 797–832.
- Kahane, Guy. 2011. "Evolutionary Dubunking Arguments." *Nous* 45 (1): 103–25.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Lemmon, E. J. 1962. "Moral Dilemmas." *Philosophical Review* 71 (2): 139–58.
- Lenman, James. 2007. "What Is Moral Inquiry?" *Proceedings of the Aristotelian Society*, suppl. 81: 63–81.
- Lenman, James. 2009. "Naturalism Without Tears." *Ratio* 22: 1–18.
- Lenman, James. 2010. "Humean Constructivism in Moral Theory." *Oxford Studies in Metaethics* 5: 175–93.
- Lenman, James. 2012. "Expressivism and Constructivism." In *Constructivism in Practical Philosophy*. Edited by James Lenman and Yonatan Shemmer, 213–25. Oxford: Oxford University Press.
- Lenman, James, and Yonatan Shemmer, eds. 2012. *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.
- Levy, Neil. 2006. "Cognitive Scientific Challenges to Morality." *Philosophical Psychology* 19 (5): 567–87.
- McGraw, A. Peter, Philip Tetlock, and Orié Kristel. 2003. "The Limits of Fungibility: Relational Schemata and the Value of Things." *Journal of Consumer Research* 30 (2): 219–29.
- Nagel, Thomas. 1979. "The Fragmentation of Value." In *Mortal Questions*. Cambridge: Cambridge University Press.
- Nussbaum, Martha. 2013. *Political Emotions: Why Love Matters for Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Quine, W. V. O. 1963. "Identity, Ostension and Hypostasis." Reprinted in *From a Logical Point of View*. New York: Harper & Row. 65–79.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge, MA: Harvard University Press.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Raz, Joseph. 2003. *The Practice of Value*. New York: Oxford University Press.
- Shemmer, Yonatan. 2012. "Constructing Coherence." In *Constructivism in Practical Philosophy*. Edited by James Lenman and Yonatan Shemmer, 159–79. Oxford: Oxford University Press.
- Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (3): 229–43.
- Stocker, Michael. 1990. *Plural and Conflicting Values*. Oxford: Oxford University Press.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109–66.

- Street, Sharon. 2008. "Constructivism About Reasons." *Oxford Studies in Metaethics* 3: 207–45.
- Street, Sharon. 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84.
- Stroop, J. Ridley. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18 (6): 643–62.
- Tessman, Lisa. 2015. *Moral Failure: On the Impossible Demands of Morality*. New York: Oxford University Press.
- Tetlock, Philip. 2003. "Thinking the Unthinkable: Values and Taboo Cognitions." *Trends in Cognitive Science* 7 (7): 320–24.
- Tetlock, Philip et al. 2000. "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals." *Journal of Personality and Social Psychology* 78 (5): 853–70.
- Walker, Margaret Urban. 1998. *Moral Understandings: A Feminist Study in Ethics*. New York: Routledge.
- Walker, Margaret Urban. 2003. *Moral Contexts*. Lanham, MD: Rowman and Littlefield.
- Williams, Bernard. 1981. "Persons, Character and Morality." In *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
- Wong, David. 2006. *Natural Moralities: A Defense of Pluralistic Relativism*. Oxford: Oxford University Press.