What’s Wrong with Morality?

C. Daniel Batson
Knoxville, Tennessee, USA

Abstract

Why do moral people so often fail to act morally? Standard scientific answers point to poor moral judgment (based on deficient character development, reason, or intuition) or to situational pressure. I consider a third possibility: a relative lack of truly moral motivation and emotion. What has been taken for moral motivation is often instead a subtle form of egoism. Recent research provides considerable evidence for moral hypocrisy—motivation to appear moral while, if possible, avoiding the cost of actually being moral—but very little evidence for moral integrity—motivation to actually be moral. The lack of truly moral motivation may, in turn, be linked to a lack of truly moral emotion, at least in response to violation of certain moral standards.

Keywords
moral emotion, moral hypocrisy, moral integrity, moral outrage

An important lesson taught by the all-too-frequent atrocities of the modern age—from the Holocaust, My Lai, Cambodia, Bosnia, Rwanda, and Darfur to corporate cover-ups of product dangers and hedge-fund fraud—is that horrendous things are not done only by monsters. People who sincerely value morality, who firmly believe they should not put their own rights and interests ahead of the parallel rights and interests of others, who think it is wrong to harm innocents, can act in ways that seem to show a blatant disregard for moral principles held dear (Arendt, 1963; Kelman & Hamilton, 1989; Lifton, 1986; Milgram, 1974; Staub, 1989; Todorov, 1991/1996). How is this possible?

Two Explanations: Bad Judgment; Situational Pressure

Scientific explanations of immoral behavior by moral people tend to be of two types. Those who approach the problem from the perspectives of social learning, character development, moral reasoning, or moral intuition are likely, one way or another, to blame bad judgment. If a person’s behavior is not adequately controlled by his or her standards, he or she must not have learned the standards well enough (Bandura, 1991; Burton & Kunce, 1995) or in the right way (Hoffman, 1977), or not have reached the necessary level of moral understanding (Kohlberg, 1976). Alternatively, this person must have a deficient moral faculty, failing to heed the intuitive voice of moral emotions (Damasio, 1994; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001, 2003; Prinz, 2006) or the universal moral grammar (Hauser, 2006; Mikhail, 2007).

Those who approach the problem from a social-influence perspective are likely to blame situational pressures—orders from a higher authority (Milgram, 1974), pluralistic ignorance (Latané & Darley, 1968), diffusion of responsibility (Darley & Latané, 1968), organizational norms (Darley, 1992), and the like. Combine these situational pressures with (a) the generality and abstractness of most moral principles (be fair; do no harm) and (b) the well-known human capacities for selective perception and rationalization, and the result is likely to be moral exclusion (Staub, 1990)—excluding certain people from those deserving moral treatment—or moral disengagement (Bandura, 1999)—deactivation of moral self-regulatory mechanisms. The problem is not bad judgment but the strength of countervailing pressures.

From a moral-judgment perspective, the remedy for moral failure is: (a) do a better job of teaching moral values and principles, (b) reason more carefully and clearly about moral matters, or (c) attend and adhere to the inner moral compass, even in the face of temptation (Bennett, 1992). A remedy is harder to prescribe from a social-influence perspective because
situational pressures are endemic to social interaction. Short of changing the structure of social institutions and environments, the most one can hope is that increased awareness of the pressures might make people less vulnerable to them—and more understanding of those who succumb.

A Third Explanation: Moral Motivation

I believe there is truth in each of these explanations of moral failure. At the same time, I believe that neither, nor the two combined, is the whole truth. Working within the framework of a general model of four forms of prosocial motivation—egoism, altruism, collectivism, and principism—each defined by a distinctive ultimate goal (own welfare, other welfare, group welfare, and upholding some moral standard or principle; see Batson, 1994, 2011), I believe we need to consider the nature of moral motivation as well. In contrast to the optimistic see Batson, 1994, 2011), I believe we need to consider the nature of moral motivation as well. In contrast to the optimistic assumption that moral individuals are motivated to act in accord with moral principles as an ultimate goal, displaying principism or moral integrity, colleagues and I have found evidence of an egoistic motive to appear moral while, if possible, avoid the cost of actually being moral. Setting aside the harsh connotations of the term, I call this subtle form of egoism moral hypocrisy. (Webster’s Desk Dictionary of the English Language, 1990, defines moral as “1. of or concerned with principles of right or wrong conduct. 2. being in accordance with such principles” [p. 589]; it defines hypocrisy as “a pretense of having desirable or publicly approved attitudes, beliefs, principles, etc., that one does not actually possess” [p. 444].)

The benefits of moral hypocrisy to oneself are obvious: one can reap the material rewards of acting selfishly and also garner the social and self-rewards of being seen and seeing oneself as upstanding and moral. Of course, if one is to garner the self-rewards, one must manage to deceive not only others but also oneself. We have found evidence of just such self-deception; unless made salient, people seem able to overlook the contradiction between their moral standards and standard-violating behavior (Batson, Thompson, Seufferling, Whitney, & Strongman, 1999).

A remedy for moral failure is especially difficult to prescribe if the failure is due to moral hypocrisy. The problem is not only that moral judgment can be bad, countervailing strong situational pressures, and many means of disengagement, but also that the goal is not to actually be moral, only to be seen by others—and to see oneself—as moral (Freud, 1930/1961; Jones & Pittman, 1982). If given enough wiggle room, seemingly good, moral people can put their own interests ahead of others’ interests without even knowing they are doing so. People who strongly endorse moral principles, and who are in relatively low-pressure situations, can fail to act morally.

Evidence of Moral Hypocrisy

To tease apart moral integrity and hypocrisy, Batson, Kobyrowicz, Dinnerstein, Kampf, and Wilson (1997) created a simple—but real—moral dilemma in the laboratory. The dilemma was a straightforward zero-sum conflict between self-interest and the interest of another person. It was simple, not complex, so there would be no problem understanding what was at stake. It was mundane and bland, not stereotypically moral (e.g., stealing or killing), to avoid scripted responses. It was a dilemma with broad consensus about the most moral course of action, not one for which opinions differed, so there would be agreement on what was right.

The dilemma involved having participants assign themselves and another participant (actually fictitious) to tasks. There was a positive-consequences task, on which each correct response earned a raffle ticket for a $30 prize, and a neutral-consequences task, on which each correct response earned nothing and which was described as rather dull and boring. One person had to be assigned to each task. Participants were told that the other person did not and would not know that they were allowed to assign the tasks.

Batson et al. (1997) used this simple dilemma to unmask the nature of moral motivation by introducing ambiguity (wiggle room) into the link between moral appearance and reality, allowing individuals to pursue self-interest without having to look selfish (Snyder, Kleck, Strenta, & Mentzer, 1979). The general strategy was, first, to make the moral standard of fairness salient and, second, to provide an opportunity to appear fair without having to incur the cost of actually being fair.

To the degree that participants are motivated to actually be fair (moral integrity), introducing ambiguity between appearance and reality should not affect the task assignment; it should still be done fairly. To the degree that participants are motivated to appear fair while, if possible, avoid the cost of being fair (moral hypocrisy), ambiguity should have an effect. Provided with wiggle room, participants should appear fair but really favor themselves. Batson et al. (1997) used two strategies to introduce ambiguity into the appearance–reality link: (a) allow participants to flip a coin in private to assign the tasks; (b) allow them to accept a prior assignment made by the experimenter.

In the absence of any cue about the morally right way to assign the tasks, Batson et al. (1997, Study 1) found that most participants (80%) gave themselves the positive-consequences task, even though in retrospect only 1 of 20 said this was the most morally right thing to do. When a fairness standard was made salient, a coin provided, and participants left alone to flip or not (Study 2), half chose to flip the coin. Of those that did not flip, 90% assigned themselves the positive-consequences task. For them, presumably, self-interested motives were stronger than moral ones. More telling, of those that did flip, 90% also assigned themselves the positive-consequences task, a significant deviation from chance. Either the coin was very charitable, or flipping it in private introduced enough ambiguity that participants could appear moral (flip the coin) yet still favor themselves (claim to have won the flip).

A self-report index of moral responsibility (which included a measure of principled moral judgment) correlated positively with flipping the coin, but it did not correlate with task assignment. This pattern of correlations provided evidence that even among those who highly endorsed moral responsibility, the dominant moral motive was hypocrisy not integrity.
Results of a third study also provided evidence for moral hypocrisy. Participants were more willing to defer to the experimenter’s impartial assignment when they knew in advance that it gave them the positive-consequences task than when it gave them the neutral one (85% versus 55%). In each of these studies, those who benefited themselves but appeared moral (by either flipping the coin or accepting the experimenter’s assignment) reported afterward that they considered their action to be quite moral.

How did participants manage to see themselves as moral yet still unfairly favor themselves? They do not seem to have distorted or disengaged their moral standards. When asked later about the most morally right way to assign the tasks, by far the most common responses were (a) give the other participant the positive-consequences task and (b) use a random method (e.g., flip a coin). Of 80 participants across the three studies, only two said that assigning oneself to the positive-consequences task was most moral. Moral standards were still there, and for many, so was the appearance of morality (e.g., flipping the coin), but real morality was rare. Evidence of moral hypocrisy rather than moral integrity has been found across a range of subsequent studies (Batson & Collins, 2011, provide a review).

**Why is Moral Integrity So Rare?**

Why is moral integrity (principlism)—motivation with the ultimate goal of upholding some moral standard or principle—not more apparent? I suspect it has to do with how we acquire these standards. Contrary to the claims of Piaget (1932) and Kohlberg (1976) that principled morality emerges as a natural product of cognitive development and adaptation to the social environment, morality is often taught. Parents are not willing to wait for their little dears to learn through experience the wisdom of principles such as fairness and justice. The child is taught to share, take turns and play by the rules, not to grab, hit, or bite.

Learned in this way, moral principles are apt to create obligations not desires, oughts not wants (Freud, 1930/1961; Heider, 1958). They are accepted as self-standards but not as part of the core sense of self. In the language of Deci, Eghrari, Patrick, and Leone (1994), the form of internalization that occurs is “introjection” rather than full integration.

Nor are attempts to encourage moral action by the administration of rewards and punishments likely to help. Theory and research on both cognitive dissonance (Festinger & Freedman, 1964) and the undermining of intrinsic motivation (Kunda & Schwartz, 1983; Lepper, 1983) provide compelling arguments that a heavy-handed carrot-and-stick approach is likely to backfire. To see oneself as acting morally due to anticipated rewards and punishments invites the inference that one does not value being moral as an ultimate goal, only as an instrumental means to obtain desired self-benefits, reinforcing the tendency to value morality extrinsically rather than intrinsically. So valued, if it is possible to obtain the self-benefits without actually being moral, as it often is, one should jump at the chance.

**Why is Moral Hypocrisy So Prevalent?**

In contrast to moral integrity, parents, educators, and religious leaders are not likely to teach hypocrisy, at least not explicitly. We are far more likely to acquire this motive through experience. Over time, we discover that just as being immoral can be costly, so can being moral. And we discover that appearance can be more important than reality.

A window through which to glimpse the emergence of moral hypocrisy is provided by observational research on the adjudication by adults of conflicts among young children, whether at home (Dunn, 1987; Dunn & Munn, 1986, 1987) or in school (Walton & Sedlak, 1982). As if in a courtroom, each child pleads his or her case before the adult—“She hit me first!” “Did not; he took my truck!” The adult then renders a verdict (Darley & Schultz, 1990). Whatever can be made to appear true to the adult, even if not actually true, decides the case. No doubt the child is shocked and confused the first time he or she, although innocent, is judged guilty because a sibling or peer was more adroit at juggling the evidence. And no doubt a key lesson is learned: adults—parents and teachers included—are not omniscient. Those who judge whether one has been good or bad, and who dispense rewards and punishments, can be deceived. Once this lesson is learned, it is only a short, natural step to convert any inclination to be moral into a desire to appear moral while, if possible, avoiding the cost of being moral. The hypocrisy motive is born.

These lessons learned in the early courtrooms of life lie deep and are a product of direct experience, not instruction. As a result, the hypocrisy motive may lie outside the realm of language, cognitive awareness, and rational control. Moreover, unlike integrity, hypocrisy has the virtue that it serves self-interest. It is no surprise, then, that much behavior that appears to be a product of moral integrity turns out, on closer inspection, to be a product of moral hypocrisy. It is also no surprise that we are prone to keep this truth even from ourselves.

**Moral Emotion**

If moral standards or principles are valued extrinsically not intrinsically, as the evidence for hypocrisy suggests, then violation of or threat to them should evoke little emotion (Batson, 2011). Such a suggestion is in sharp conflict with recent claims about the prevalence and power of moral emotion (Greene et al., 2001; Haidt, 2001, 2003; Prinz, 2006). Still, I think we need to consider the possibility that even though violation of moral standards can evoke strong emotion, much of this emotion is not evoked by the moral violation per se.

Haidt (2003) suggested that moral emotions are “those emotions that are linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent” (p. 853). This definition strikes me as too broad, including as it does fear of terrorism, joy at an enemy’s failure or a friend’s success, and so on. I think we need a more focused definition. Extending to moral motivation the functional framework that
has illuminated the relation between valued state, emotion, motivation, and behavior for other prosocial motives (Batson, 1994, 2011), it may be more useful to identify as moral those emotions evoked by violation of or threat to some personally valued moral standard, principle, or ideal. Such violations or threats might, for example, evoke anger, disgust, or contempt, as claimed by Rozin, Lowery, Imada, and Haidt (1999). They might also evoke concern, sorrow, sadness, or distress. If so, these emotions should evoke truly moral motivation to uphold the threatened standard (i.e., moral integrity). There may be positive moral emotions as well, such as joy at witnessing the upholding of a valued moral standard, but as they do not directly evoke moral motivation, I shall not consider them here.

Using this more focused, functional criterion, many emotions routinely thought moral need reexamination. For example, to the extent that guilt and shame are evoked by threat to the concept of oneself as good and moral (or at least not immoral; Janoff-Bulman, Sheikh, & Hepp, 2009) rather than threat to the standard per se, they are not truly moral emotions (contrary to the classification offered by Haidt, 2003; also see Tangney, Stuewig, & Mashek, 2007). Similarly, to the extent that empathic concern, sympathy, and compassion are evoked by threat to the welfare of a cared-for other, they are not truly moral emotions. Each of these sets of emotions is capable of producing strong motivation—egoistic motivation to repair or maintain one’s self-image (Tangney & Dearing, 2002; Tangney et al., 2007); altruistic motivation to increase the welfare of the cared-for other (Batson, 1991, 2011). These motives can, in turn, produce behavior judged moral (i.e., in accord with moral standards). But as long as upholding principle is an unintended consequence and not an ultimate goal, this behavior is not morally motivated.

When considering the prevalence and power of moral emotion, it may help to distinguish two broad domains of morality: conflict and propriety. Conflict standards address the consideration a person should give to the interests of others in situations in which their interests conflict with one’s own. Fairness and justice principles, as well as principles proscribing harm, form the core of conflict morality. Properly standards address the natural and social order—the way things should be. To say that it is immoral to defile nature, break the law, violate sexual taboos, disobey or disrespect authority, or act unpatriotically is to invoke propriety principles. Although these two domains have overlapping boundaries (e.g., many people think harming innocents violates the natural order), our experience of them seems different.

Propriety morality is acquired primarily by modeling and habit, not instruction. It is usually learned early and well, as simply the way things are. The child has little need to resist or dissemble. As a result, propriety standards are likely to be valued intrinsically, and their violation is capable of producing both strong motivation and strong emotion—especially disgust and anger (Haidt, 2003). Most of us do not find it difficult to uphold our propriety principles; we find it difficult even to imagine breaking them. The research by Haidt (2001) on “moral dumbfounding” and by Tetlock, Kristel, Elson, Green, and Lerner (2000) on “taboo tradeoffs” focuses on violations of propriety standards, standards so thoroughly internalized and “natural” that their validity and necessity seem intuitively obvious (e.g., avoid sexual intercourse with a sibling; don’t sell body organs).

Conflict standards are less intuitively obvious. As suggested earlier, they are learned through instruction and not fully internalized. We are not dumbfounded or shocked to learn that someone has acted unfairly—or that we might be tempted to do so. My suspicion about the scarcity of truly moral emotion is directed at the emotion evoked by violation of conflict standards, not propriety standards. Absence or weakness of such emotions could explain the absence or weakness of truly moral motivation noted earlier.

**Emotion Evoked by Violation of Conflict Standards: The Case of Moral Outrage**

Consider moral outrage, which has been defined as anger evoked by the violation of or threat to principles of justice, fairness, and not doing harm (Batson et al., 2007; Darley & Pittman, 2003; Hoffman, 2000; Montada & Schneider, 1989). Colleagues and I have explored the possibility that much of the anger or outrage evoked by violation of such principles is not really moral outrage (i.e., anger at the violation or threat to the principles per se). Specifically, we have distinguished moral outrage from two nonmoral forms of anger that can occur when someone is intentionally and unfairly harmed: personal anger—anger at undeserved harm to oneself or someone with whom one identifies (Montada & Schneider, 1989)—and empathic anger—anger at undeserved harm to a cared-for other (Hoffman, 2000).

These three emotional reactions to violation of standards proscribing undeserved harm may all be expressed—even experienced—as anger, indignation, and outrage. Still, from an appraisal-theory perspective on emotion (Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1987), these forms of anger are importantly, if subtly, different because they are evoked by different aspects of the situation. As a result, they produce different motivational consequences. The appraisal producing moral outrage is that the standard has been violated, a moral wrong has occurred. The resulting motive is to reestablish the standard, a form of moral integrity (principlism) or truly moral motivation. The appraisal evoking personal anger is that I have been undeservedly harmed. The resulting motive is to undo the harm and/or punish the harm-doer, forms of self-protective or self-assertive egoistic motivation. The appraisal evoking empathic anger is that a person for whom I care has been harmed. The resulting motive is, again, to undo the harm and/or punish the harm-doer, forms of other-protective altruistic motivation.

Neither personal anger nor empathic anger is a truly moral emotion or source of truly moral motivation. These forms of anger may motivate action described as moral (e.g., exacting retribution), but the goal is not to uphold moral standard; it is to protect one’s own or the cared-for other’s interests. These three forms of anger could, of course, co-occur.
Anger at unfair advantage. Moral outrage, personal anger, and empathic anger cannot be distinguished by self-reports alone. All three are likely to lead to reports of anger. However, it is possible to distinguish among them by experimentally manipulating conditions relevant to the different appraisals and observing the resulting pattern of self-reported anger. The relevant conditions are unfair treatment of self, of a cared-for other, and of an unknown other. If exactly the same unfair act occurs in each condition, moral anger or outrage should appear in all three, whereas personal anger should appear only in the first, empathic anger only in the second.

Conducting two experiments that included these three conditions, Batson et al. (2007) found evidence of personal anger and empathic anger, but not of moral outrage. Participants perceived unfair treatment of another—even a stranger for whom they had not been induced to care—to be as unfair and immoral as they perceived unfair treatment of themselves. But in the absence of empathic concern for the other, unfair treatment of another evoked little anger. Anger was triggered by unfair treatment of self or a cared-for other, not by unfairness per se.

Anger at torture. Batson, Chao, and Givens (2009) extended this research to include identity-based personal anger. They presented U.S. university students with a news story (actually fictitious) reporting a case of physical torture to extract military information. The torture involved severe beating, waterboarding, and ripping off fingernails with pliers. For half of the participants, the story told of a U.S. Marine tortured by al-Qaida in Iraq; for the other half, the story was identical except that a Sri Lankan soldier was tortured by Tamil rebels. Participants were then asked to report (a) their emotional response, including feelings of anger, and (b) whether the torture was morally wrong.

Participants considered the torture described in the news story to be clearly morally wrong regardless of who was tortured—U.S. Marine or Sri Lankan soldier. But there was a strong effect on anger for who was tortured. Torture of a U.S. Marine, with whom participants shared national identity, produced considerable anger. The same torture of a Sri Lankan soldier, with whom participants had no identity ties, produced little anger. Greater anger at torture of a U.S. Marine than of a Sri Lankan soldier was evidence of identity-based personal anger, not moral outrage at torture per se.

Anger at unfair exclusion. In a third attempt to find evidence of moral outrage, O’Mara, Jackson, Batson, and Gaertner (2011) looked at anger following unfair social exclusion. They found that although the unfair exclusion was judged to be morally wrong in both cases, participants reported more anger after they were unfairly excluded than after a stranger was unfairly excluded. Unfair exclusion of the stranger produced very little anger. Moreover, when given a subsequent opportunity to punish the instigator of unfair exclusion, punishment was high only when the self was unfairly excluded, not the stranger. This pattern of punishment suggested that participants were motivated by personal revenge rather than by defense of principle. In sum, each of the four experiments to date that have been designed to look for moral outrage unconfounded with personal and empathic anger following violation of conflict standards has failed to find any clear evidence of moral outrage.

Why All the Talk of Moral Outrage?

Results of these experiments pose a puzzle: Why do people—psychologists included—speak so readily and persistently of moral outrage when clear evidence of it is so hard to find? The results certainly do not rule out the possibility that moral outrage exists, but they do suggest that it is not nearly as prevalent or powerful as has been assumed.

Talk of moral outrage rather than personal (or empathic) anger may be no accident. I suspect it reflects a motivated attempt at moral engagement by victims of harm (as distinct from moral disengagement by perpetrators of harm; Bandura, 1999). By calling personal or empathic anger moral outrage, the victim (or person caring for the victim) sends a message to three distinct audiences. To oneself one says: my anger and whatever behavior it evokes are good and right, not personal and petty. To the harm-doer one says: by inflicting this harm, you have violated standards and shown yourself to be morally deficient; there will be hell to pay. To bystanders one says: there has been a moral offense; you should join in and help right this wrong.

These messages transform anger triggered by harm to self or a cared-for other into a righteous crusade. No wonder such talk.

Of What Use Is Morality?

Moral principles and standards are often assumed to be intrinsically motivating and “designed to regulate behavior” (Prinz, 2006, p. 41). Yet the research I have briefly reviewed casts doubt on this optimistic view. Rather than intrinsically motivating, we may adhere to principle only to avoid social and self-punishments or gain social and self-rewards, including the rewards of being seen by others and ourselves as good and moral.

If moral principles rarely motivate and guide our behavior, why do we have them? Perhaps they serve to evaluate and control others’ behavior, holding their feet to the fire of censure and approbation even as they, like we, seek to wriggle free by substituting the appearance of morality for the reality (DeScioli & Kurzban, 2009). From this admittedly cynical perspective, conflict morality may be the language of victims; propriety morality, the language of the status quo.

Conclusion

Problems arising from poor moral judgment are real and important, as are problems caused by situational pressures. But I think neither fully explains what’s wrong with morality. There are also problems with the prevalence and power of moral motivation and emotion. If we wish to encourage moral people to act morally, I doubt we can rely on moral judgment, moral motivation, and moral emotion alone. We also need to tap nonmoral resources.
References


Bandura, A. (1999). Moral disengagement in the perpetration of inhumani-


Haidt, J. (2001). The emotional dog and its rational tail: A social intu-


Kohlberg, L. (1976). Moral stages and moralization: The cognitive-


Latane, B., & Darley, J. M. (1968). Group inhibition of bystander interven-


