

# Moral Hypocrisy: Appearing Moral to Oneself Without Being So

C. Daniel Batson, Elizabeth R. Thompson, Greg Seufferling, Heather Whitney, and Jon A. Strongman  
University of Kansas

How can people appear moral to themselves when they fail to act morally? Two self-deception strategies were considered: (a) misperceive one's behavior as moral and (b) avoid comparing one's behavior with moral standards. In Studies 1 and 2 the authors documented the importance of the 2nd strategy but not the 1st. Among participants who flipped a coin to assign themselves and another participant "fairly" to tasks, even a clearly labeled coin that prevented misperception did not produce a fair result (Study 1). Inducing behavior–standard comparison through self-awareness did (Study 2). Study 3 qualified the self-awareness effect: When moral standards were not salient before acting, self-awareness no longer increased alignment of behavior with standards. Instead, it increased alignment of standards with behavior and produced less moral action. Overall, results showed 3 different faces of moral hypocrisy.

Moral people often fail to act morally. One of the most important lessons to be learned from the tragically common atrocities of the modern age—the Holocaust, My Lai, Cambodia, Bosnia, Rwanda, corporate cover-ups of product dangers—is that horrendous things are not done only by monsters. People who sincerely value morality, who firmly believe that they should not put their own rights and interests ahead of the parallel rights and interests of others, can act in ways that seem to show a blatant disregard for the moral principles they hold dear (Arendt, 1963; Bandura, 1991; Darley, 1992; Kelman & Hamilton, 1989; Lifton, 1986; Milgram, 1974; Staub, 1989; Todorov, 1996). How is this possible?

## Two Explanations: Learning Deficits and Situational Pressures

Social psychological explanations of immoral action by moral people tend to be of two types. First, those who approach the problem from a social learning perspective are likely to blame a learning deficit. If a person's behavior is not adequately controlled by his or her standards, then the standards must not have been learned well enough or in the right way (Bandura, 1977; Burton & Kuncle, 1995; Hoffman, 1977; Wright, 1971).

Second, those who approach the problem from a social influence perspective are likely to blame situational pressures: orders from a higher authority (Milgram, 1974), conformity pressure (Asch, 1956), foot-in-the-door processes (Freedman & Fraser, 1966), pluralistic ignorance (Latané & Darley, 1968), diffusion of responsibility (Darley & Latané, 1968), and the like. Combine these situational pressures with (a) the generality and abstractness of

most moral principles (be fair; do no harm; do unto others . . .) 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, more generally, *moral disengagement* (Bandura, 1991)—deactivation of moral self-regulatory mechanisms in specific situations.

From a social learning perspective, the remedy for moral failure is for society to do a better job of teaching moral values and principles. People need standards strong enough to guide behavior, 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, especially in hierarchical organizations. The most that one can hope is that increased awareness of the pressures might make people less vulnerable to them and more understanding of those who succumb. As Bandura (1990) summarized,

Mechanisms of moral disengagement operate not only in the perpetration of inhumanities under extraordinary circumstances, but in everyday situations where people routinely perform activities that bring personal benefits at injurious costs to others. Given the many psychological devices for disengagement of moral control, societies cannot rely solely on individuals, however honorable their standards, to provide safeguards against inhumanities. (p. 27)

## A Third Explanation: Moral Hypocrisy

We believe that there is truth in each of these explanations of moral failure. At the same time, we believe that neither one, nor the two combined, is the whole truth. Batson, Kobryniewicz, Dinnerstein, Kampf, and Wilson (1997) suggested that one needs 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 *moral integrity*, Batson et al. found evidence of a motive to appear moral in one's own and others' eyes while, if possible, avoiding the cost of actually being moral. Setting aside the harsh connotations of the term, they called this motivation *moral hypocrisy*. (*Webster's Desk Dictionary of the English Language*, 1990, defines *moral* as "1. of or concerned with principles of right or wrong

---

C. Daniel Batson, Elizabeth R. Thompson, Greg Seufferling, Heather Whitney, and Jon A. Strongman, Department of Psychology, University of Kansas.

Thanks to Javette Hayes and Patricia Schoenrade of William Jewell College for their contributions to pilot research on moral hypocrisy and self-awareness and to Nadia Ahmad, David Lishner, Peter Sherrard, and Jo-Ann Tsang for helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to C. Daniel Batson, Department of Psychology, University of Kansas, Lawrence, Kansas 66045-0001.

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 to oneself of moral hypocrisy 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.

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 motivation is weak, countervailing situational pressures strong, and means of disengagement many, but also that the goal is not to actually be moral, only to see oneself and to be seen by others as moral (Freud, 1930/1961; Jones & Pittman, 1982). From this perspective, seemingly good, moral people will put their own interests ahead of the interests of others if they can appear not to do so. Even people who strongly endorse moral principles and who are in relatively low-pressure situations can fail to act morally.

### *Evidence of Moral Hypocrisy*

To examine the nature of moral motivation, Batson et al. (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 that individuals would have no problem understanding what was at stake. It was mundane and bland, not stereotypically moral and dramatic (e.g., an issue of life or death), so as to get less scripted responses. It was a dilemma for which there would be consensus about the morally right course of action, not one for which opinions would conflict, so participants would know what action was likely to be considered moral.

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 the participant a raffle ticket for a \$30 prize, and a neutral consequences task, on which each correct response earned the participant 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 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 into the link between moral appearance and moral 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 either provide or not provide an opportunity to appear fair without having to incur the cost of actually being fair.

To the degree that participants were motivated to uphold moral principle (moral integrity), introducing ambiguity between appearance and reality should not affect task assignment because to appear fair without really being fair would not uphold principle. To the degree that participants were motivated to appear moral while, if possible, avoiding the cost of being moral (moral hypocrisy), ambiguity should affect task assignment. It should lead participants to appear fair but really favor themselves. Only when the appearance–reality link is unambiguous should moral hypocrisy produce a moral result. Batson et al. (1997) used two strategies to introduce ambiguity into the appearance–reality link: They al-

lowed participants to flip a coin to assign the tasks (Study 2), and they allowed participants to accept a prior assignment made by the experimenter (Study 3).

In the absence of any cue about the morally right way to assign the tasks, Batson et al. (1997) found that few participants (20%) assigned the other person to the positive consequences task; most gave themselves the positive consequences task, even though in retrospect only 1 of 20 said this was the most morally right thing to do (Study 1). When a fairness standard was made salient, a coin was provided, and participants were left alone to flip the coin or not (Study 2), half chose to flip the coin. Of those that did not flip the coin, only 10% assigned the other person to the positive consequences task. For them, presumably, self-interest motives were stronger than moral ones. More telling, of those that did flip the coin, only 10% assigned the other person to the positive consequences task, a significant deviation from chance. Either the coin was very charitable, or, with self-interest an issue, flipping the coin introduced enough ambiguity that participants could appear moral (flip the coin) while still really favoring themselves (claim to have won the flip). What is more, although a self-report index of moral responsibility correlated positively with flipping the coin, it did not correlate with task assignment. This pattern of results provided evidence that the dominant moral motive operating in this situation 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 it favored them; 85% accepted the experimenter's assignment when it gave them the positive consequences task compared with only 55% when it gave them the neutral consequences task. In each of these studies, those who benefited themselves but appeared moral (by either flipping the coin or accepting the experimenter's assignment) reported that they considered their action to be highly moral.

How did participants in the Batson et al. (1997) research manage to negotiate this dilemma so that they could 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 to (a) give the other participant the positive consequences task or (b) use a random method (e.g., flip a coin). Of the 80 participants across the three studies, only 2 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 (the outcome of the coin flip was biased).

### *The Present Research: How Is Moral Hypocrisy Possible?*

The goal of the present research was to go beyond a demonstration of the phenomenon of moral hypocrisy to consider how it is possible. At issue is not simply appearing moral to others (Jones & Pittman, 1982). If one is to gain the self-rewards for being moral and, more important perhaps, to avoid the self-punishments for being a hypocrite, then one must appear moral to oneself. How does a person manage to appear moral to oneself while violating one's moral standards to serve self-interest?

One possible way is to engage in self-deception. Biologist Robert Trivers (1985, pp. 415–420; see also Alexander, 1987, pp.

114–125) suggested that people have evolved self-deception strategies that allow them to avoid perceiving discrepancies between their self-serving actions and their moral standards. Baumeister and Newman (1994) also suggested that self-deception is necessary for phenomena like moral hypocrisy to work. Responding to Kunda's (1990) call for further analysis of motivated reasoning—cognitive processes mobilized in the service of personal desires—Baumeister and Newman portrayed the motivated individual as an intuitive lawyer, who selectively gathers evidence and then reasons from it to make a case for a predetermined, particular conclusion. They contrasted this metaphor with that of the intuitive scientist (Fiske & Taylor, 1984), who sets aside preconceptions and dispassionately, if fallibly, seeks an accurate conclusion. Baumeister and Newman (1994) reasoned that self-deception must be integral to the thinking of the intuitive lawyer: "To reach a preferred conclusion, self-regulation must serve the ends of self-deception, but self-deception fails if it is acknowledged too openly" (p. 15; see also Bersoff, 1999).

Given that self-deception is a concept that has been used widely and with varying meanings in both philosophy and psychology, it is important to consider (a) what form of self-deception is implied by the motive of moral hypocrisy and (b) what strategies might serve this motive. Sometimes self-deception has been thought to require that a person simultaneously hold two contradictory beliefs but be unaware of one (e.g., Gur & Sackheim, 1979) or that a person believe what at some level he or she knows is not so (e.g., Demos, 1960). Such paradoxical forms of self-deception are not required for moral hypocrisy. To reach the goal of appearing moral to oneself, it is not necessary to believe one is moral while being unaware of also believing one is immoral, or to believe one is moral while at some level knowing one is not. It is sufficient that one engages in what philosopher Alfred Mele (1987) calls "ordinary self-deception" or "desire-influenced manipulation of data" (p. 126). The goal of moral hypocrisy can be reached if one can manipulate the data so as to avoid confronting the discrepancy between one's self-serving behavior and one's moral standards.

How can one manipulate the data to avoid this discrepancy? Assuming that (a) one has behaved in a way that violates his or her moral standards and (b) responsibility for the behavior cannot be denied or justified, ordinary self-deception strategies that would serve moral hypocrisy can be classified as one of two types. First, one could perceive the behavior as moral (i.e., as being in line with his or her moral standards), even though it actually is not. Second, one could avoid comparing the behavior with his or her moral standards. The first of these two strategies, if available, seems preferable because the second leaves one vulnerable to anything that might focus attention on the behavior–standard discrepancy. If a person can convince himself or herself that he or she has acted morally, then the behavior can be scrutinized from the perspective of his or her standards with impunity, even pride. The first strategy fails only if the self-serving behavior cannot be perceived as being in line with one's moral standards.

Might participants in the Batson et al. (1997) studies have been able to perceive their self-serving behavior as being in line with their standards? For example, might they have been able to perceive their coin flip as fair even when it was biased in their favor? It certainly seems possible. Participants were left alone to decide whether to flip the coin and, if they did flip it, how to interpret the outcome. It was not specified who should be assigned to which

task if the coin came up heads or if it came up tails. This was left up to the participants. Further, participants were not required to make their specification explicit before they flipped the coin.

In a way analogous to the perversion of hypothesis-testing research that Kerr (1998) has called *HARKing* (hypothesizing after the results are known), participants may have failed to clearly specify the consequences of heads and tails in advance. Once they saw the result of the flip was heads (or tails), they may have found it relatively easy to "remember" that this result assigned them to the positive consequences task. They could then enjoy both the positive consequences task and the knowledge that they had assigned the tasks fairly because they had flipped the coin. As Batson et al. (1997) proposed when interpreting their results,

With self-interest an issue, flipping the coin introduced enough ambiguity into the decision process that participants could feel moral while still favoring themselves. ("It's heads. Let's see, that means . . . I get the positive task." "It's tails. Let's see, that means . . . the other participant gets the neutral task.") (p. 1342)

Participants could not lose—and they almost always did not.

### Study 1: Labeling the Coin

To the degree that participants in the Batson et al. (1997) research used this first self-deception strategy, it should be relatively easy to diminish if not eliminate the moral hypocrisy effect. All that is required is to specify in advance who gets which task if the coin comes up heads or tails. Accordingly, in Study 1 we placed participants in exactly the same moral dilemma encountered by participants in Batson et al.'s Study 2. We made the moral standard of fairness salient, suggested a coin flip, and provided a coin. On each side of the coin, however, there was a sticker; one side said "SELF to POS" and the other said "OTHER to POS." We reasoned that this labeled coin should make it more difficult, if not impossible, to perceive that one had acted fairly and morally should one take the positive consequences task after a flip produced "OTHER to POS." If, when using this labeled coin, the hypocrisy effect disappeared (i.e., there was no longer a significant deviation from 50% in the direction of assigning the self to the positive consequences task), then we would have evidence the previously observed hypocrisy effect was a product of the first self-deception strategy. If the effect remained, then we would have to look for an explanation elsewhere.

### Method

**Participants.** Participants for Study 1 were 40 introductory psychology students (20 women, 20 men) at the University of Kansas. They received credit toward a course requirement for participating. On the basis of probes during debriefing, 1 additional woman was dropped from the design and replaced because she expressed doubt about the presence of another participant.

**Procedure.** When participants arrived for their appointment, they were greeted by a same-gender experimenter, who escorted them to a research cubicle on another floor of the building. The experimenter explained that 2 participants were taking part in the study and it was important that they not talk to or even see each other, so they were being met at different locations. (In reality, there was no other participant.)

**Task consequences.** Once seated in the research cubicle, participants were given a written introduction to the study and left alone to read it. The

introduction explained that the study was part of a research project on "how a variety of task characteristics affect people's feelings and reactions after performing a task":

We are currently examining the effects of task consequences on feelings and reactions. Task consequences can be either positive, negative, or neutral. In this particular study, we are focusing on the effects of *positive* task consequences—where correct responses are rewarded, and there are no consequences for incorrect responses. As a comparison, we are also considering *neutral* task consequences—where feedback is given about correct and incorrect responses, but there are no consequences at all.

The introduction explained that 1 of the 2 participants in the session would be assigned to do a positive consequences task and the other to do a neutral consequences task: "By having different types of tasks performed simultaneously by different people, we are able to control for the subtle effects on feelings and reactions of time of day, day of week, and so on." The introduction stated that the 2 participants would work on their tasks independently and would never meet. Finally, the introduction explained the consequences:

To ensure that you care about the consequence of your performance, it is necessary to use real consequences. The consequences we are using in this study are the following: If you are assigned to the task with *positive consequences*, then for an incorrect response you will receive nothing. For each correct response, however, you will receive one raffle ticket; the prize in the raffle is a \$30.00 gift certificate at the store of your choice. Only participants in this study are eligible for this raffle, so if you receive a number of tickets you have a good chance of winning a gift certificate.

If you are assigned to the task with *neutral consequences*, then you will simply perform a task and be given feedback about your performance. Your responses are not related to any consequences. (You should be aware that most participants assigned to the neutral consequences task find it rather dull and boring. If you are assigned to this task, we hope you will not find it too bad.)

*The chance to assign the tasks using a labeled coin.* After participants finished reading the introduction, the experimenter returned, answered any questions, asked participants to name the store from which they would like a gift certificate were they to win the raffle, and gave them a folder containing a typewritten information sheet describing the task assignment. Participants were left alone to read this sheet, which began,

There is one aspect of the procedure of this study that we purposely did not explain earlier, but about which we can tell you now. In addition to studying the effect of task consequences—positive or neutral—on feelings and reactions, we are also interested in studying the effect of assigning people to different tasks. To this end, in some sessions one of the two participants is asked to assign both participants to their tasks. The other participant is entirely unaware of this, simply being told that the tasks were assigned by chance.

As you have probably guessed by now, *you* are the participant in this session who will assign the tasks. You must assign one of you—yourself or the other participant—to the positive consequences task and the other of you to the neutral consequences task.

Most participants feel that giving both people an equal chance—by, for example, flipping a coin—is the fairest way to assign themselves and the other participant to the tasks (we have provided a coin for you to flip if you wish, designating one side "SELF to POS" for positive consequences for yourself and the other side "OTHER to POS" for positive consequences for the other participant). But the decision is entirely up to you. You can assign yourself and the other participant

however you choose. The other participant does not and will not know that you are assigning tasks; he or she will think that the task assignment was purely by chance. Because of this and because the two of you will never meet, your anonymity is assured.

The first sentence of the preceding paragraph made the moral standard of fairness salient, and the statement in parentheses explained the labels on the coin.

Taped to the inside of the folder that contained the information sheet was a clear plastic pouch. The pouch held a coin (a quarter) for participants to flip if they wished. A round sticker covered each side. One side was labeled "SELF to POS," and the other was labeled "OTHER to POS."

Finally, the sheet instructed participants to indicate their decision on an enclosed task assignment form and to return the form to the folder. On the form, one line read, "Participant assigned to *positive* consequences task \_\_\_\_"; the next line read, "Participant assigned to *neutral* consequences task \_\_\_\_." Instructions on this form read as follows: "Please indicate your assignment of yourself and the other participant to the tasks by putting an *S* in one blank (for self) and an *O* in the other blank (for the other participant). Thank you." The response on this form was the major dependent measure.

*Perceptions of the morality of one's decision.* Once participants filled out the form and indicated that they were ready to proceed, the experimenter returned, collected the folder, gave participants several questionnaires concerning the task-assignment decision, and left them alone to complete the questionnaires while the experimenter ostensibly went to prepare the tasks. Among the questions were two that assessed perceptions of morality. The first, which was open-ended, asked about the morally right way to assign the tasks: "In your opinion, what was the most morally right way to assign the task consequences?" Three blank lines were provided to write an answer. Later, in a different section, participants were asked to rate the morality of their own decision on a 9-point scale: "Do you think the way you made the task assignment was morally right?" (1 = *not at all*, 9 = *yes, totally*). The first question assessed participants' standard for the morally right way to assign the tasks; the second assessed perceptions of the morality of their assignment behavior.

*Debriefing.* When participants finished these questionnaires, the experimenter returned and asked about their reactions to the study so far. This question led into a full and careful debriefing in which the true purpose of the research and all deceptions were explained. Participants did not actually perform a task. Because they had been offered the opportunity to earn raffle tickets, all participants were informed that they would receive 20 tickets in a raffle to be held after the completion of data collection. Participants were then thanked for their time and excused. Once data collection was complete, the raffle was held and a \$30 gift certificate was awarded to the winner.

## Results

*Task assignment.* Roughly two thirds (28) of our 40 participants chose to flip the labeled coin (cf. half, 10 of 20, who flipped the unlabeled coin in Batson et al., 1997, Study 2). Of the 12 who chose not to flip the coin, 10 assigned themselves to the positive consequences task, leaving the dull and boring task for the other participant; only 2 (.17) assigned the other person to the positive consequences task. More importantly, of the 28 who chose to flip the coin, only 4 (.14) assigned the other person to the positive consequences task. This proportion differed significantly from the .50 that would occur by chance when flipping a coin ( $z = 3.78$ ,  $p < .001$ ), indicating that we had once again found evidence of moral hypocrisy. There were no reliable gender effects (main effects or interactions) on these or any other measures. (For convenience, all statistical tests are reported two-tailed, even for directional predictions.)

Many of those who flipped the coin seemed to share our belief that having labels on the coin made it more difficult to misperceive a biased flip as fair. Asked during debriefing whether they thought it would make any difference if the coin had no labels, over half of the participants explicitly said that without labels it would be easier to cheat (e.g., "It would be easier to pick me if the coin was not labeled. You could fool yourself and say, 'No, I meant that side'"; "Without the labels you'd be more likely to change your mind after the flip"; "People would probably cheat their way out and take the positive for themselves"; "It's not as concrete; you could fudge it"; "You could play mind games until you came to the conclusion that you get the positive"). Still, labeling the coin did not eliminate—or even noticeably reduce—the moral hypocrisy effect.

*Most morally right way to assign the tasks.* When participants were asked in the open-ended question about the most morally right way to assign the tasks, the vast majority agreed with the standard made salient on the introduction sheet. Of the 40 participants, 31 said that flipping the coin (or using some other random method) was the most morally right way to assign the tasks. This was the most frequent response among both those who flipped the coin and those who did not. Of the 28 who flipped the coin, 24 said flipping the coin was the most morally right way, 2 said assigning the other participant to the positive consequences task was the most morally right way, and 2 said that there was no morally right way to assign the tasks. Of the 12 who did not flip the coin, 7 said flipping the coin was the most morally right way, 3 said assigning the other participant to the positive consequences task was the most morally right way, and 2 said assigning oneself to the positive consequences task was the most morally right way. Clearly, for many participants there was a discrepancy between task-assignment behavior and moral standard.

*Perceived morality of the assignment.* When asked later about the morality of the way they assigned the tasks, participants who flipped the coin rated what they had done more morally right ( $M = 7.64$  on the 9-point scale) than did participants who did not flip the coin ( $M = 4.67$ ),  $t(38) = 4.04$ ,  $p < .0005$ . This was true even though those who flipped were as likely as those who did not flip to assign themselves the positive consequences task. The ratings of perceived morality were very close to those found by Batson et al. (1997; in Study 2,  $M_s = 7.30$  and  $4.00$  for those who did and did not flip the coin, respectively) and suggest that those who flipped the coin felt they had acted quite morally. Those who did not flip the coin felt they had acted less morally, but even they, on average, rated the morality of their action near the midpoint of the scale. Of those who assigned themselves to the positive consequences task, those who did so after flipping the coin felt that they had acted highly morally ( $M = 7.42$ ), whereas those who did so without flipping the coin did not ( $M = 3.90$ ),  $t(32) = 4.66$ ,  $p < .0005$ .

## Discussion

Clearly, we were wrong to think that making it harder to perceive a biased coin flip as fair would eliminate the moral hypocrisy effect. Labeling the coin did nothing to increase the fairness of the task assignment among those who flipped it. Apparently, if moral hypocrisy in the task-assignment procedure is a product of self-deception, then the self-deception is not of the form

suggested by Batson et al. (1997), whereby participants took advantage of not having to call the coin in advance to convince themselves that regardless how the coin came up, it was in their favor. As had Batson et al., we found that most people said the most morally right way to assign the tasks was to flip the coin (or to use some other random method); only 2 of 40 (.05) said that assigning themselves to the positive consequences task was the most morally right way. Still, in keeping with the findings of Batson et al., we found that the vast majority, 34 of 40 (.85), assigned themselves to the positive consequences task. Even with the coin labeled, participants who flipped a coin assigned themselves to the positive consequences task at a rate well beyond the limits of chance ( $p < .001$ ).

It seems almost certain that some participants who flipped the coin found themselves faced with a coin that said "OTHER to POS" yet still assigned themselves to the positive consequences task. This behavior is difficult to explain by the first self-deception strategy we described: perceiving one's self-serving behavior as moral. The results of Study 1 do not prove that this strategy is not used to promote moral hypocrisy, but they do suggest that this strategy is not necessary.

Accordingly, in Study 2 we turned our attention to the second strategy: avoiding the comparison of one's immoral behavior (unfairly assigning oneself to the positive consequences task) with one's moral standards (being fair). How might this strategy be used in the task-assignment situation? Faced with the statement making salient the standard of fairness, those who flipped the coin but went against the result to favor themselves might have reasoned to the point "I flipped the coin, which is fair" and stopped, without adding "but I ignored the result, which is not." They may have avoided assessing their behavior in light of the salient moral standard.

## Study 2: Looking at Moral Hypocrisy in a Mirror

In Study 1 we used a labeled coin to make it more difficult for participants to use the first self-deception strategy; in Study 2 we sought to make it more difficult for participants to use the second strategy. We did this by making some participants self-aware.

Self-awareness manipulations, such as looking at one's face in a mirror, have been found to heighten awareness of discrepancies between behavior and salient personal standards, creating pressure to act in accord with standards (Wicklund, 1975). In the moral domain, for example, Diener and Wallbom (1976) induced self-awareness in research participants who had the opportunity to cheat on a test. They found that those who were self-aware cheated significantly less often than those who were not self-aware.

If people who are motivated to appear moral yet avoid the cost of actually being moral rely on not comparing their self-serving behavior with their moral standards, then introducing self-awareness should reduce or eliminate the hypocrisy effect. It should make the behavior–standard discrepancy salient, creating pressure to act in accord with the standard. Pursuing this logic, in Study 2 we had participants face exactly the same moral dilemma and read the same statement about fairness used in Study 1. The only procedural differences were, first, that the coin provided for participants to flip if they wished was not labeled and, second, a mirror was present. Participants in the high self-awareness condition were seated so they were facing the mirror while deciding how

to assign the tasks; those in the low self-awareness condition were also seated facing the mirror, but it was turned to the wall so they saw only its back.

In the low self-awareness condition, we expected to replicate the Batson et al. (1997, Study 2) moral hypocrisy effect, as we had in Study 1. That is, we expected a majority of participants, even of those who flipped the coin, to assign themselves to the positive consequences task. If an important contributor to the hypocrisy effect is the ability to avoid or ignore the discrepancy between one's behavior and one's moral standards (the second self-deception strategy), then in the high self-awareness condition, in which participants are face-to-face with themselves and the discrepancy, the cost of appearing fair without being fair should increase, making a biased flip difficult, if not impossible. As a result, among participants who flip the coin in this condition, the hypocrisy effect should disappear; assignment of the self to the positive consequences task should be roughly 50%.

## Method

**Participants.** Participants for Study 2 were 52 introductory psychology students (32 women, 20 men) at the University of Kansas. They received credit toward a course requirement. Using a randomized block procedure, we assigned half (16 women, 10 men) to each experimental condition (low self-awareness, high self-awareness). On the basis of probes during debriefing, 3 additional students (1 woman and 2 men, all in the low self-awareness condition) were dropped from the design and replaced because they expressed doubt about the presence of another participant.

More women were included in the sample because two of the three experimenters were women, and we wished to keep the gender of the participant and the experimenter the same to minimize cross-gender self-presentation concerns (Jones & Pittman, 1982). The different numbers of men and women did not seem to be a problem because, as in Study 1, there were no reliable gender effects (main effects or interactions) on any of the reported measures.

**Procedure.** With the exception of removing the labels on the coin, removing all mention of the labels in the written instructions, and introducing the self-awareness manipulation, we kept the procedure and measures for Study 2 identical to those for Study 1. Therefore, only the self-awareness manipulation will be described.

**Self-awareness manipulation.** As participants were escorted into the research cubicle, the experimenter noted and apologized for the presence of a mirror (91.5 cm wide  $\times$  51 cm high), explaining that the cubicle was being shared with a research group doing a perception study. The mirror was leaning against the wall, propped on the only table in the room, at which participants were asked to take a seat. When seated, the mirror was directly in front of participants, roughly 60 cm away.

In the low self-awareness condition, the mirror was turned to the wall, so participants saw only its gray back. In the high self-awareness condition, the mirror was facing them, so participants saw themselves. A sign in the corner of the mirror said "ANDERSON STUDY—DON'T TOUCH PLEASE," and participants were asked not to move the mirror. None did.

## Results

**Task assignment.** Approximately half (23) of the 52 participants in Study 2 chose to flip the coin. Separating those who flipped the coin from those who did not, Table 1 presents the proportion of participants in each experimental condition who assigned the other person to the positive consequences task. In the low self-awareness condition, results closely replicated the findings of Batson et al. (1997) and of Study 1: Of the 13 who chose

Table 1

*Proportion of Participants Assigning the Other Person to the Positive Consequences Task and Mean Rated Morality of Assignment, Study 2*

Did participant flip coin?	Self-awareness condition	
	Low	High
No		
Proportion	.15	.38
Rated morality	4.77	5.19
<i>n</i>	13	16
Yes		
Proportion	.15	.50
Rated morality	8.00	8.40
<i>n</i>	13	10

*Note.* For the proportion measure, assignment of oneself to the positive consequences task was coded 0; assignment of the other participant to the positive consequences task was coded 1. Morality was rated on a 9-point scale in response to the item "Do you think the way you made the task assignment was morally right?" (1 = *not at all*, 9 = *yes, totally*).

not to flip the coin, 11 (.85) assigned themselves to the positive consequences task, leaving the dull and boring task for the other participant; only 2 (.15) assigned the other person to the positive consequences task. More importantly, of the 13 who chose to flip the coin, again only 2 (.15) assigned the other person to the positive consequences task. This proportion differed significantly from .50 ( $z = 2.50$ ,  $p < .02$ ), indicating that we had, once again, found evidence of moral hypocrisy.

Results in the high self-awareness condition, in which participants made the task-assignment decision while facing themselves in the mirror, were quite different. Here, the coin was scrupulously fair; 10 of the 26 participants chose to flip the coin, and of these, 5 assigned the other participant and 5 assigned themselves to the positive consequences task. Six of the 16 who chose not to flip the coin (.38) assigned the other participant to the positive consequences task. A 2 (flip: no, yes)  $\times$  2 (self-awareness: low, high) loglinear analysis (Fienberg, 1980; Wickens, 1989) produced a significant self-awareness main effect,  $\chi^2(1, N = 52) = 4.55$ ,  $p < .04$ , indicating an overall tendency for less discrepancy between behavior and the salient fairness standard in the high self-awareness condition than in the low self-awareness condition. Neither the main effect for whether participants flipped the coin nor the interaction approached significance, both  $\chi^2s < 1$ .

**Perceived morality of the assignment.** Table 1 also includes mean ratings by participants of the degree to which they thought the way they made the task assignment was morally right. As can be seen, in spite of the effect of the self-awareness manipulation on task assignment, the mirror had very little effect on perceived morality ( $M_s = 6.38$  and 6.42 on the 9-point scale for the low and high self-awareness conditions, respectively),  $F(1, 48) = 0.48$ ,  $ns$ . There was, however, a strong effect of whether participants acted in accord with the moral standard of fairness by flipping the coin ( $M_s = 5.00$  and 8.17 for those who did not and did flip the coin, respectively),  $F(1, 48) = 29.31$ ,  $p < .0005$ . The interaction did not approach significance ( $F < 1$ ).

Adding task assignment (self to positive consequences task vs. other participant to positive consequences task) as a variable to

create a  $2$  (coin flip)  $\times 2$  (self-awareness)  $\times 2$  (task assignment) design, there was again a highly significant main effect on rated task morality for whether participants flipped the coin,  $F(1, 44) = 55.93, p < .0005$ ; there was also a highly significant main effect for task assignment ( $M_s = 5.65$  and  $8.27$  for those who assigned the self and the other participant to the positive consequences task, respectively),  $F(1, 44) = 31.37, p < .0005$ . These two main effects were qualified by a highly significant and easily interpretable Coin Flip  $\times$  Task Assignment interaction,  $F(1, 44) = 25.23, p < .0005$ . Rated morality was high for participants who assigned the other to the positive consequences task regardless of whether they flipped the coin; it was also high for participants who assigned themselves to the positive consequences task if they flipped the coin ( $M_s$  ranged from  $8.13$  to  $8.29$ ). Rated morality was low only for those who assigned themselves to the positive consequences task and did not flip the coin ( $M = 3.76$ ).

Among those assigning themselves to the positive consequences task, the difference between those who did and did not flip the coin was highly significant not only in the high self-awareness condition, where the coin flip seemed quite fair,  $t(13) = 5.57, p < .0001$  ( $M_s = 8.80$  and  $3.50$ , respectively), but also in the low self-awareness condition, where it did not,  $t(20) = 6.53, p < .0001$  ( $M_s = 7.82$  and  $4.00$ , respectively). Flipping the coin, even if one adjusted the result to benefit oneself, was sufficient to merit self-ratings of high morality as long as one was not directly confronted with the behavior–standard discrepancy.

*Most morally right way to assign the tasks.* The difference in task assignment between the low and high self-awareness conditions was not due to a difference in perceptions of the most morally right way to assign the tasks. As can be seen in Table 2, when participants were asked in the open-ended question about the most morally right way to assign the task, a majority in each self-awareness condition agreed with the standard made salient on the introduction sheet. Of the 52 participants, 32 said that flipping the coin (or use of some other random method) was the most morally right way. Fifteen of the remaining 20 respondents said that assigning the other participant to the positive consequences task was the most morally right way; 4 said that there was no morally right way to make the assignment; and 1 said that assigning oneself to the positive consequences was the most morally right way. The difference in this pattern between conditions did not approach statistical significance,  $\chi^2(3, N = 52) < 2.00, ns$ .

Table 2  
*Participants' Reports of Most Moral Way  
to Assign Tasks, Study 2*

Most morally right way to assign tasks	Self-awareness condition		Total
	Low	High	
Other participant to positive task	6	9	15
Random method (flip coin)	17	15	32
Self to positive task	1	0	1
No morally right way	2	2	4
Total	26	26	52

## Discussion

To determine whether the second self-deception strategy—avoiding the comparison of one's behavior and moral standards—was an important contributor to moral hypocrisy, we used a mirror manipulation of self-awareness. We found that self-awareness had little effect on the procedure participants used to make the task-assignment decision. In the low self-awareness condition, half, 13 of 26 (.50), flipped the coin, whereas in the high self-awareness condition, only slightly fewer, 10 of 26 (.38), did. This difference was not reliable. Nor did self-awareness change participants' reports of the most morally right way to make the decision. The majority in each self-awareness condition said that flipping the coin (or some other random method) was the most moral way; most of the rest said that assigning the other person to the positive consequences task was the most moral way.

Self-awareness did, however, significantly affect the actual assignment. Among participants in the low self-awareness condition who flipped the coin, only 2 of 13 participants (.15) assigned the other participant to the positive consequences task. This significant deviation from chance in the direction of self-interest replicated the results of Batson et al. (1997) and of Study 1. As before, the 11 participants in the low self-awareness condition who flipped the coin and assigned themselves to the positive consequences task rated their action as highly moral. It is almost certain that some of these 11 were practicing moral hypocrisy; they appeared moral and rated themselves as such, yet they avoided the cost of actually being moral.

*Self-awareness eliminates moral hypocrisy.* We replicated the moral hypocrisy effect in the low self-awareness condition but not in the high self-awareness condition. In front of the mirror, the coin became scrupulously fair; half of the participants assigned themselves and half assigned the other participant to the positive consequences task. This difference across the self-awareness conditions was precisely the pattern expected if making the behavior–standard discrepancy salient rendered the second self-deception strategy ineffective and forced participants to actually be moral in order to appear moral.

*Those who did not flip the coin.* There was little evidence of moral hypocrisy among participants who did not flip the coin. This was not because there was evidence of moral integrity; it was because there was little evidence of even the appearance of morality. In the low self-awareness condition, the vast majority (11 of 13) who did not flip the coin assigned themselves to the positive consequences task, admitting that this was not the most morally right way to make the assignment and that what they had done was not moral. Even in the high self-awareness condition, a majority of those who did not flip the coin (10 of 16) chose to assign themselves to the positive consequences task, admitting what they had done was not moral. If not totally comfortable with the discrepancy between their behavior and standards, those who chose not to flip the coin were able to live with it even in the high self-awareness condition, but they made no claim to have acted morally.

*Does self-awareness stimulate moral integrity?* The coin flip in the high self-awareness condition seemed quite fair; participants who flipped won exactly half of the time. Why? We have suggested that self-awareness rendered the second self-deception strategy ineffective and forced those wishing to appear fair to actually be fair. This explanation assumes that the dominant mo-



tive under high self-awareness was still moral hypocrisy. However, there is another, less cynical possibility. High self-awareness may have stimulated motivation to uphold the moral principle of fairness as an ultimate goal, evoking a desire actually to be fair, that is, moral integrity.

To test the validity of these two very different accounts of the effect of self-awareness on moral behavior in Study 2, we conducted a third study. In this study we removed the opportunity to mask hypocrisy with a coin flip, and we manipulated both the salience of a moral standard and self-awareness. Gibbons and Wicklund (1982) had found that in the presence of a salient moral standard, self-awareness increased helping with a moral cause, whereas in the absence of a salient moral standard, self-awareness actually reduced helping. Duval and Lalwani (in press) recently proposed an intriguing extension of self-awareness theory to account for these and similar results.

Duval and Lalwani (in press) pointed out that changes of two types can enable a person to reduce a perceived discrepancy between one's behavior and standards. The person can either bring behavior in line with standards, as has often been observed in past self-awareness studies, or bring standards in line with behavior: "Both changing self toward standards and changing standards toward self appear to be equally viable ways to reduce the negative affect caused by self-standard discrepancy" (Duval & Lalwani, in press, p. 5). Duval and Lalwani presented evidence supporting this proposal in a performance context. They also suggested that a key determinant of which change will occur is the point in time at which relevant standards are made salient. If standards are made salient prior to the opportunity to behave, then behavior is likely to be moved toward standards; if they are not, then standards are likely to be moved toward behavior.

### Study 3: Self-Awareness and Integrity or Hypocrisy?

In Study 3 we built on the Duval and Lalwani (in press) logic to determine the nature of the moral motivation operating when people are self-aware. Is it moral integrity or moral hypocrisy? We used the same task-assignment dilemma used in the first two studies but without any mention of flipping a coin. Instead, we manipulated moral standard salience (low, high) and self-awareness (low, high).

Moral integrity means that one's behavior is directed toward upholding one's moral standards; standards are not simply changed to fit behavioral inclination. If self-awareness stimulates moral integrity, then increased self-awareness should produce unidirectional pressure to bring behavior in line with standards. The accommodation under high self-awareness should be of behavior to standard, even in a low standard-salience condition. Accordingly, a moral integrity account predicted a main effect of self-awareness in Study 3; there should be increased assignment of the other participant to the positive consequences task in the high self-awareness condition for participants in both standard-salience conditions. Moreover, in the low-standard-salience/high-self-awareness condition, reports of the most morally right way to assign the tasks should not shift away from concern for the other participant's rights and interests toward self-benefit.

In contrast, a moral hypocrisy account predicted an interaction between standard salience and self-awareness. If, prior to the assignment decision, a moral standard is made salient, then in-

creased self-awareness should make it more difficult to appear moral yet fail to act in accord with this salient standard. As a result, in the high standard-salience condition, high self-awareness should lead to increased assignment of the other participant to the positive consequences task. However, if no standard is mentioned, then it should be possible to reduce the behavior-standard discrepancy by bringing moral standards in line with behavior. As a result, in the low standard-salience condition, high self-awareness should lead to no increase—possibly even a decrease—in assignment of the other participant to the positive consequences task. Additionally, in the low-standard-salience/high-self-awareness condition, reports of the most morally right way to assign the tasks should shift toward self-benefit.

### Method

**Participants.** Participants for Study 3 were 56 introductory psychology students (40 women, 16 men) at the University of Kansas. They received credit toward a course requirement for participating. Using a randomized block procedure, we assigned 14 participants (10 women, 4 men) to each condition of the 2 (low standard salience, high standard salience)  $\times$  2 (low self-awareness, high self-awareness) design. It was not necessary to exclude any participants for suspicion.

Experimenters were the same as for Study 2, so once again, more women were included in the sample to keep the gender of the participant and the experimenter the same and so minimize cross-gender self-presentation concerns. As before, the different number of men and women did not seem to be a problem because there were no reliable gender effects (main effects or interactions) on any of the reported measures.

**Procedure.** The procedure for Study 3 was exactly the same as for Study 2, except for three changes: (a) The statement about "giving both people an equal chance—by, for example, flipping a coin" was removed from the task-assignment information sheet, (b) no coin was provided, and (c) the salience of the moral standard was manipulated. Self-awareness was manipulated exactly as in Study 2, and measures were the same as in Studies 1 and 2. Therefore, only the standard-salience manipulation will be described.

**Manipulation of standard salience.** In the low standard-salience condition, after explaining that participants would be assigning "one of you—yourself or the other participant—to the positive consequences task and the other of you to the neutral consequences task," the information sheet made no mention of the most morally right way to make the assignment. The sheet simply continued: "The decision is entirely up to you." In the high standard-salience condition, the explanation of task assignment was followed by the observation:

Most participants feel that giving the other person the positive consequences task is the most morally right way to assign themselves and the other participant to the tasks. But the decision is entirely up to you.

In both standard-salience conditions, the rest of the information sheet was the same as in the previous studies.

### Results

**Task assignment.** Table 3 presents the proportion of participants in each experimental condition of Study 3 who assigned the other person to the positive consequences task. The low-standard-salience/low-self-awareness condition was a conceptual replication of the procedure of Batson et al.'s (1997) Study 1. Therefore, it is reassuring to note that the proportion assigning the other person to the positive consequences task in this condition, 4 of 14 (.29), was similar to the proportion in their study, 4 of 20 (.20).



Table 3

*Proportion of Participants Assigning the Other Person to the Positive Consequences Task and Mean Rated Morality of Assignment, Study 3*

Salience of moral standard condition	Self-awareness condition	
	Low	High
Low		
Proportion	.29	.00
Rated morality	6.14	6.50
High		
Proportion	.29	.57
Rated morality	5.14	6.29

*Note.* In each experimental condition,  $n = 14$  (10 women, 4 men). For the proportion measure, assignment of oneself to the positive consequences task was coded 0; assignment of the other participant to the positive consequences task was coded 1. Morality was rated on a 9-point scale in response to the item "Do you think the way you made the task assignment was morally right?" (1 = *not at all*, 9 = *yes, totally*).

Going beyond replication, we found in the high-standard-salience/low-self-awareness condition that making a moral standard salient in the absence of self-awareness had no effect on the decision; again, 4 of 14 (.29) assigned the other participant to the positive consequences task.

As predicted by the moral hypocrisy account, self-awareness increased the proportion of participants assigning the other person to the positive consequences task when the statement about the morally right way to assign the tasks was included in the information sheet (to .57) but decreased it when the statement was omitted (to .00). The difference in the high self-awareness condition between those who were presented with a prior standard and those who were not was found to be significant using a loglinear analysis ( $z = 3.06, p < .005$ ), as was the factorial interaction,  $\chi^2(1, N = 56) = 5.17, p < .025$ . There was no evidence of the self-awareness main effect predicted by a moral integrity account,  $\chi^2 = 0.00, ns$ . Because proportions were the same in both low self-awareness conditions, the standard-salience main effect matched the interaction. The clear standard-salience effect in the high self-awareness condition and the significant interaction were entirely consistent with a moral hypocrisy account but inconsistent with a moral integrity account.

*Perceived morality of the assignment.* Table 3 also includes mean ratings by participants of the degree to which they thought the way they made the task assignment was morally right. There

were no reliable main effects, nor was there a reliable interaction on these ratings, all  $F_s < 1.30, p_s > .25$ . There were, however, strong effects when task assignment (self to positive consequences task vs. other participant to positive consequences task) was added as a variable to create a 2 (standard salience)  $\times$  2 (self-awareness)  $\times$  2 (task assignment) design. As in Study 2, participants who assigned the other person to the positive consequences task rated the morality of their action as being high regardless of the experimental condition ( $M_s$  ranged from 7.75 to 8.63 on the 9-point scale). However, ratings by participants who assigned themselves to the positive consequences task varied by condition. Those in the high standard-salience condition rated their standard-discrepant behavior as less moral ( $M = 3.75$ ) than did those in the low standard-salience condition ( $M = 6.04$ ),  $F(1, 36) = 11.14, p < .002$ . This difference was especially pronounced in the high self-awareness condition ( $M_s = 3.17$  and 6.50, for high and low standard salience, respectively,  $t(18) = 3.51, p < .002$ ), although the Standard Salience  $\times$  Self-Awareness interaction was not significant,  $F(1, 36) = 2.30, p < .15$ .

*Most morally right way to assign the tasks.* Participants' open-ended reports of the most morally right way to assign the tasks provided further support for a moral hypocrisy account. Following the Duval and Lalwani (in press) suggestion, this account predicted that if a standard is made salient prior to behavior, then self-awareness will lead one to adjust behavior toward the standard, whereas if a standard is not made salient prior to the behavior, then self-awareness will lead one to adjust standards toward behavior.

As can be seen in Table 4, participants in the low-standard-salience/low-self-awareness condition, who showed little concern about acting in accord with moral principles, did not seem very troubled by a behavior-standard discrepancy. Only 4 of 14 participants (.29) assigned the other person to the positive consequences task, yet 10 of 14 (.71) reported that the most morally right way to assign the tasks was either to give the other person the positive consequences task (7) or to use a random method (3). In the two high standard-salience conditions, in which participants had read that "most participants feel that giving the other person the positive consequences task is the most morally right way," a majority agreed that this was the most moral way (21 of 28). Agreement was especially high in the high-standard-salience/high-self-awareness condition, where 13 of 14 participants said this was the most morally right way to assign the tasks.

Participants in the low-standard-salience/high-self-awareness condition responded quite differently from those in the other three conditions. Only 3 of the 14 participants in this condition said that

Table 4

*Participants' Reports of Most Moral Way to Assign Tasks, Study 3*

Most morally right way to assign tasks	Low standard salience		High standard salience		Total
	Low self-awareness	High self-awareness	Low self-awareness	High self-awareness	
Other participant to positive task	7	3	8	13	31
Random method (flip coin)	3	2	3	0	8
Self to positive task	0	4	0	0	4
No morally right way	4	5	3	1	13
Total	14	14	14	14	56

assigning the other person to the positive consequences task was the most morally right way; 4 said that assigning themselves to the positive consequences was the most morally right way, a response not given in any other condition in Study 3; and 5 said that there was no morally right way to assign the tasks. The difference across conditions was well beyond what would be expected by chance,  $\chi^2(9, N = 56) = 24.24, p < .005$ .

To simplify and clarify this difference, we coded responses in the first two rows of Table 4 as other benefiting and responses in the last two rows as self-benefiting. Using this simplified classification, in the low self-awareness conditions (including both low and high standard salience), 21 of 28 participants said that an other-benefiting way was most moral (10 of 14 and 11 of 14 in the low and high standard-salience conditions, respectively). In the high-standard-salience/high-self-awareness condition, 13 of 14 participants said an other-benefiting way was most moral, but in the low-standard-salience/high self-awareness condition, only 5 of 14 participants said this. A loglinear analysis of the difference between the two high self-awareness cells was highly significant ( $z = 2.67, p < .01$ ), producing a marginal Standard Salience  $\times$  Self-Awareness interaction,  $\chi^2(1, N = 56) = 3.55, p < .06$ . It seemed clear that in the low-standard-salience/high-self-awareness condition, in which all participants assigned themselves to the positive consequences task, the moral standard moved toward behavior, whereas in the high-standard-salience/high-self-awareness condition, behavior moved toward the standard.

## Discussion

In Study 3 we sought to see still more of the nature of moral motivation. Unlike our Study 2 and Batson et al.'s (1997) Study 2, in which a moral standard was made salient and ambiguity was introduced to allow participants to appear moral yet still serve self-interest, we manipulated whether a moral standard was salient and provided no ambiguity. We did this to clarify the nature of the motivation that led to the fair and unbiased result of the coin flip in the high self-awareness condition of Study 2. Was it because self-awareness stimulated moral integrity or because it produced another form of moral hypocrisy?

Results of Study 3 supported moral hypocrisy, not moral integrity, predictions. There was an increase in the proportion of participants assigning the other person to the positive consequences task in the high self-awareness condition compared with the low self-awareness condition when the standard was salient, but there was a decrease when the standard was not salient. Moreover, among high self-awareness participants who assigned themselves to the positive consequences task, those in the low standard-salience condition rated their action as being moral, whereas those in the high standard-salience condition did not. As might be expected, given the results on rated morality, most participants in the low-standard-salience/high-self-awareness condition reported either that the most morally right way to assign the tasks was to give themselves the positive consequences task or that there was no morally right way to assign the tasks, justifying self-benefit. These responses were quite different not only from responses in the other three conditions of Study 3 but also from responses in Studies 1 and 2 and in the Batson et al. (1997) studies, where even without making a standard salient almost no one said that giving themselves the positive consequences task was most moral, in spite

of most having done just that. If self-aware individuals are motivated by moral integrity, then they should not shift their standards in the direction of self-interest to deal with a behavior–standard discrepancy; they should bring behavior in line with standards. Clearly, this is not what happened in the low-standard-salience/high-self-awareness condition of Study 3.

## General Discussion

In our attempt to understand how moral hypocrisy is possible, we considered the role of two strategies for pursuing what Mele (1987) called “ordinary self-deception.” In Study 1, we considered the possibility that participants faced with the task-assignment dilemma might have been able to convince themselves that they had fairly won the coin flip by waiting to specify the consequences of heads and tails until after the result of a flip was known. We found that specifying the consequences in advance by labeling the coin did not diminish the moral hypocrisy effect. This finding suggests that the effect was not dependent on the first self-deception strategy.

In Study 2 we used a self-awareness manipulation to assess the role of a second self-deception strategy: avoiding or ignoring the discrepancy between one's behavior and one's moral standards. We found that having participants assign the tasks while facing themselves in a mirror eliminated the moral hypocrisy effect. In contrast to the biased coin flip in the low self-awareness condition, participants who flipped the coin in the high self-awareness condition of Study 2 were scrupulously fair. The other person got the positive consequences task half of the time, and the participants got it half of the time. Elimination of the hypocrisy effect under high self-awareness is precisely what would be expected if hypocrisy is made possible by the second self-deception strategy.

We suggested earlier that the second self-deception strategy is not ideal because it leaves one vulnerable. How, precisely, do people manage to avoid confronting moral behavior–standard discrepancies? Our research did not directly address this question, but we can propose two possibilities. The first possibility is that people use affective alarms to suspend cognition. In discussing the role of self-deception in the intuitive lawyer's efforts to reach a predetermined outcome, Baumeister and Newman (1994) suggested that one key way self-deception occurs is through the use of affective cues that warn one of thoughts not to think. By attending to these cues, the source of which is unanalyzed, thought can remain strategically incomplete. The intuitive lawyer avoids reaching undesirable conclusions and having to face their implications.

We suspect that some of those practicing moral hypocrisy in our studies were in a state of suspended cognition, warned by prescient affective cues (intuitions, perhaps?) not to attend to the behavior–standard discrepancy. This state could not be maintained, however, when facing oneself—and the discrepancy—in the mirror.

Second, at least some people may have a relatively high tolerance for behavior–standard discrepancy in their moral lives. The responses of participants in Studies 1 and 2 who did not flip the coin underscore this possibility. Most gave themselves the positive consequences task, even though they said that the most morally right way to assign tasks was either to give the other person the positive consequences task or to flip the coin; most also rated the morality of what they had done as being relatively low. These people did not appear particularly troubled by the blatant discrep-

ancy between their behavior and their standard, and they seemed quite willing to admit the relatively low morality of their action. They showed no signs of either self-deception or moral hypocrisy; they frankly acknowledged a moral lapse.

High tolerance for a behavior–standard discrepancy may have existed also among those who engaged in moral hypocrisy. Some participants may have recognized and tolerated the discrepancy between the apparent fairness of flipping the coin and their knowledge that they rigged the outcome. If so, theirs was a bounded tolerance; it disappeared with high self-awareness. Facing the appearance–reality discrepancy in the mirror was, apparently, too much to bear.

### *Three Faces of Moral Hypocrisy*

Moral hypocrisy—the motivation to appear moral while, if possible, avoiding the cost of actually being moral—is not easy to see. On the surface, it looks like moral integrity—the motivation to uphold moral principle as an ultimate goal. To differentiate between moral hypocrisy and moral integrity, it is necessary to look beneath the surface to see whether the appearance of morality is only a mask for an underlying self-serving motive. In our experiments we were able to catch at least a glimpse of three different faces of moral hypocrisy.

*A familiar face.* The first had been seen before by Batson et al. (1997). Participants who flipped a coin in Study 1 and in the low self-awareness condition of Study 2 appeared to be upholding the moral principle of fairness, which had been made salient. Yet, the biased outcome of the coin flip showed that at least some of the participants did not really allow the coin flip to determine the task assignment; they made sure the coin gave them the positive consequences task. The ambiguity that the coin introduced into the appearance–reality link allowed them to appear moral yet still serve self-interest at the expense of the welfare of the other participant.

Batson et al. (1997) only studied women when they first unmasked this face of moral hypocrisy. By including both men and women in Studies 1 and 2, replicating the hypocrisy effect, and finding no gender effects, we confirmed that moral hypocrisy is not gender specific. The desire to appear moral yet still serve self-interest exists among men as well as women.

*A frightening new face.* We were able to see two new faces of moral hypocrisy in Study 3. One appeared when participants saw themselves in the mirror in the low standard-salience condition. These participants dealt with the potential discrepancy between giving themselves the positive consequences task and doing what was morally right in a very interesting way. Consistent with Duval and Lalwani's (in press) suggestion, they did not bring their behavior in line with standards; they brought their standards in line with behavior. All 14 participants gave themselves the positive consequences task, and 9 of these said either that this was the most moral thing to do (4) or that there was no morally right way to assign the tasks, making morality irrelevant (5); only 5 said that some form of other-oriented action was most moral. This perception of the most morally right way to assign the tasks was dramatically different from the perception of participants in all other conditions of our studies and in Batson et al.'s (1997) studies. With this perception, most participants in the low-standard-salience/high-self-awareness condition could and did feel that they had

acted morally yet still serve self-interest.

Seeing this face of moral hypocrisy was frightening for two reasons. First, although occasions on which a person is asked to make a moral decision in front of a mirror are rare, those on which a person is asked to do so while self-aware—because others are watching, because he or she feels accountable for his or her action, because someone challenges his or her action, and so on—are quite common. Often on these occasions, relevant moral standards are not stated in advance. Therefore, it seems that many everyday moral decisions, including many contributing to the horrendous moral failures of our age such as the Holocaust, occur in the conceptual equivalent of our low-standard-salience/high-self-awareness condition (see Todorov, 1996), the condition in which we saw the lowest rate of moral outcome.

The second reason for concern is that a less moral outcome is not the only effect of moral hypocrisy under these circumstances. Our results suggest that there is likely to be a transformation of moral standards as well. The transformation can involve, our participants' responses suggest, either redefining self-interest as moral or redefining moral decisions as not moral. Stimulated by self-awareness to reduce the behavior–standard discrepancy by bringing standards in line with self-oriented behavior, one may create new standards that justify—even encourage—self-interested behavior in the future. What is more, attempts to challenge, expose, and monitor morally questionable behavior may only backfire; they are likely to heighten self-awareness, which may accelerate the process. The result can be dramatic. As Todorov (1996) said of Nazi death camp guards:

Guards who committed atrocities never stopped distinguishing between good and evil. Their moral faculty had not withered away. They simply believed the "atrocities" was in fact a good thing and thus not an atrocity at all. (p. 129)

*Glimpse of a third face.* The third face of moral hypocrisy was the hardest to see. We caught only an indirect glimpse of it by comparing responses of high self-awareness participants for whom a relevant moral standard had been made salient prior to the task-assignment decision (Studies 2 and 3) with responses of those for whom a standard had not been made salient (Study 3). When a standard had been made salient, self-awareness led to no erosion of the standard; it led to an increased alignment of behavior with the standard, producing a more moral outcome. Although this raised the possibility that self-awareness had stimulated moral integrity, a closer look suggested that self-awareness simply made it harder to live with a behavior–standard discrepancy, which led those motivated by moral hypocrisy to be moral as the least costly way to appear moral.

Two pieces of evidence, admittedly indirect, supported this conclusion. Both were found in the low-standard-salience/high-self-awareness condition of Study 3. First, moral behavior was lowest in this condition, yet the rated morality of the behavior was highest (Table 3). Second, perceptions of the most morally right way to assign the tasks changed to accommodate self-interest (Table 4). Each of these findings seems difficult to explain if self-awareness stimulates moral integrity, but each easily fits a moral hypocrisy account. Therefore, we believe that even under high self-awareness, what we are seeing is not moral integrity but another face of moral hypocrisy. Moral hypocrisy can, it seems, assume many guises.

### *Links to Two Other Perspectives on Moral Motivation*

*Moral hypocrisy and the egoism of the altruistic personality.* Recognition of the many guises of moral hypocrisy may shed light on the motivation underlying what has been called the altruistic personality. Several researchers have suggested that certain personality measures, including measures of (a) personal and social responsibility, (b) a positive view of human nature, and (c) self-reported concern for the welfare of others tap a stable personality dimension that leads one to care for others (Oliner & Oliner, 1988; Rushton, 1980; Staub, 1974). Batson, Bolen, Cross, and Neuringer-Benefiel (1986) challenged this conclusion. They found that elements of this altruistic personality cluster correlated positively with helping a young woman in distress by taking shocks in her stead, but the correlations were found only among those who thought they would have to watch her continued suffering if they did not help. Among those who thought they would not have to watch, the correlations vanished.

Batson et al. (1986) argued this pattern of correlations indicated that the motivation associated with the altruistic personality was egoistic, not altruistic. Rather than being directed toward the ultimate goal of relieving the young woman's suffering, responses of those high on this personality dimension seemed directed toward the goal of not having to face an uncomfortable discrepancy between their failure to help and their self-image as good, moral people. When it was relatively difficult to escape the discrepancy, those high on the altruistic personality dimension helped more; when it was easy, they did not.

The present analysis of moral hypocrisy suggests a more precise understanding of the Batson et al. (1986) results. It seems likely that for those who score relatively high on measures of the altruistic personality, moral standards are chronically accessible and salient. It also seems likely that knowing one must sit and watch the young woman's continued suffering, which is caused by one's own unwillingness to help, would heighten awareness of a behavior-standard discrepancy. If these two conceptual links are valid, then the Batson et al. (1986) results map directly onto the results of our Study 3: a strong effect for standard salience in the high self-awareness condition and no effect in the low self-awareness condition. The parallel is striking and leads us to propose that the egoistic motivation that characterizes the so-called altruistic personality is moral hypocrisy. Those scoring high on this personality dimension are motivated to appear moral while, if possible, avoiding the cost of actually being moral. The many faces and guises of moral hypocrisy may account for why sometimes the altruistic personality seems to be stimulating moral integrity (Eisenberg et al., 1989; Staub, 1974). We do not wish to claim that moral integrity does not exist (even in the situations we have studied), only that appearances can be so deceiving that the prevalence and power of moral integrity has almost certainly been overestimated.

*Moral hypocrisy and moral disengagement.* In his impressive and insightful analysis of the mechanisms of moral disengagement, Bandura (1991) identified a range of psychological resources available to an individual motivated to appear moral yet still serve self-interest. The mechanisms include (a) reconstruing possibly reprehensible conduct (by means of moral justification, palliative comparison, and euphemistic labeling); (b) displacing and diffusing responsibility; (c) minimizing, ignoring, and misconstruing

detrimental consequences; and (d) dehumanizing and blaming victims. Bandura and others (e.g., Darley, 1992; Kelman & Hamilton, 1989; Lifton, 1986; Milgram, 1974; Staub, 1989) have suggested that mechanisms such as these help explain human moral disasters from the Holocaust to corporate cover-ups, as well as the less dramatic but more common moral failures of everyday life.

Bandura (1991) believed that these mechanisms are called on when situational inducements lead self-interest to overpower moral control; they serve to disengage moral motivation so one can pursue self-interest with equanimity. If, however, the ultimate goal of moral motivation is to appear moral while, if possible, avoiding the cost of actually being moral (moral hypocrisy), then rather than using these mechanisms to disengage moral motivation, an individual may use them in the service of it. The mechanisms may do more than allow the individual not to appear and feel immoral: they may allow him or her to appear and feel moral. In the masquerade of moral hypocrisy, moral rationalization and moral motivation are allies, not opponents.

We know no better illustration of this alliance than the ironic first pages of Jane Austen's *Sense and Sensibility* (1811/1995). The well-off Mr. John Dashwood and his wife are discussing how he should fulfill the promise to his dying father to "do everything in his power" (p. 3) to care for his stepmother and three stepsisters, whose financial situation is now dire. "When he gave his promise to his father, he meditated within himself to increase the fortunes of his sisters by the present of a thousand pounds apiece" (p. 3) from the seven thousand pounds he was to inherit.

However, soon he sees his moral responsibility more clearly. His wife reminds him of their son: "'How could he answer it to himself to rob his child, and his only child too, of so large a sum?'" (p. 6). He cuts the imagined annuity in half, to five hundred pounds apiece, affirming, "'I would not wish to do anything mean. . . . One had rather on such occasions do too much than too little'" (p. 7). Thinking further, he abandons the idea of an annuity altogether. "'Whatever I may give them occasionally will be of far greater assistance than a yearly allowance, because they would only enlarge their style of living if they felt sure of a larger income and would not be sixpence the richer for it at the end of the year'" (p. 9). His wife goes even further: "'I am convinced within myself that your father had no idea of your giving them any money at all. . . . They will be much more able to give you something'" (p. 9). At last, Mr. Dashwood settles on a plan that will "'strictly fulfill my engagement'" (p. 10):

"When my mother removes into another house [he is evicting her from the one that has been her home] my services shall be readily given to accommodate her as far as I can. Some little present of furniture too may be acceptable then. . . ." He finally resolved that it would be absolutely unnecessary, if not highly indecorous, to do more for the widow and children of his father. (p. 10)

Such is the human capacity to appear moral—even to oneself—without being so.

### References

- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Arendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. New York: Viking.

- Asch, S. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, 70 (9, Whole No. 416).
- Austen, J. (1995). *Sense and sensibility*. New York: Doherty. (Original work published 1811)
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1990). Selective activation and disengagement of moral control. *Journal of Social Issues*, 46(1), 27-46.
- Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development: Vol. 1. Theory* (pp. 45-103). Hillsdale, NJ: Erlbaum.
- Batson, C. D., Bolen, M. H., Cross, J. A., & Neuringer-Benefiel, H. (1986). Where is the altruism in the altruistic personality? *Journal of Personality and Social Psychology*, 50, 212-220.
- Batson, C. D., Kobryniewicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335-1348.
- Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin*, 20, 3-19.
- Bennett, W. J. (1992). *The de-valuing of America: The fight for our culture and our children*. Colorado Springs, CO: Focus on the Family.
- Bersoff, D. M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin*, 25, 28-39.
- Burton, R. V., & Kunce, L. (1995). Behavioral models of moral development: A brief history and integration. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Moral development: An introduction* (pp. 141-171). Boston: Allyn & Bacon.
- Darley, J. M. (1992). Social organization for the production of evil. *Psychological Inquiry*, 3, 199-218.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Demos, R. (1960). Lying to oneself. *Journal of Philosophy*, 57, 588-595.
- Diener, E., & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 10, 107-111.
- Duval, S. D., & Lalwani, N. (in press). Objective self-awareness and causal attributions for self-standard discrepancies: Changing self or changing standards of correctness. *Personality and Social Psychology Bulletin*.
- Eisenberg, N., Miller, P. A., Schaller, M., Fabes, R. A., Fultz, J., Shell, R., & Shea, C. L. (1989). The role of sympathy and altruistic personality traits in helping: A reexamination. *Journal of Personality*, 57, 41-67.
- Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: MIT Press.
- Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. Reading, MA: Addison-Wesley.
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4, 195-202.
- Freud, S. (1961). *Civilization and its discontents* (J. Strachey, Trans.). New York: Norton. (Original work published 1930)
- Gibbons, F. X., & Wicklund, R. A. (1982). Self-focused attention and helping behavior. *Journal of Personality and Social Psychology*, 43, 462-474.
- Gur, R. C., & Sackheim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37, 147-169.
- Hoffman, M. L. (1977). Moral internalization: Current theory and research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 86-135). New York: Academic Press.
- Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (pp. 231-262). Hillsdale, NJ: Erlbaum.
- Kelman, H. C., & Hamilton, V. C. (1989). *Crimes of obedience: Toward a social psychology of authority and responsibility*. New Haven, CT: Yale University Press.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196-217.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention. *Journal of Personality and Social Psychology*, 10, 215-221.
- Lifton, R. J. (1986). *The Nazi doctors: Medical killing and the psychology of genocide*. New York: Basic Books.
- Mele, A. R. (1987). *Irrationality: An essay on akrasia, self-deception, and self-control*. New York: Oxford University Press.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.
- Oliner, S. P., & Oliner, P. M. (1988). *The altruistic personality: Rescuers of Jews in Nazi Europe*. New York: Free Press.
- Rushton, J. P. (1980). *Altruism, socialization, and society*. Englewood Cliffs, NJ: Prentice Hall.
- Snyder, M. L., Kleck, R. E., Strenta, A., & Mentzer, S. J. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology*, 37, 2297-2306.
- Staub, E. (1974). Helping a distressed person: Social, personality, and stimulus determinants. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 293-341). New York: Academic Press.
- Staub, E. (1989). *The roots of evil: The origins of genocide and other group violence*. New York: Cambridge University Press.
- Staub, E. (1990). Moral exclusion, personal goal theory, and extreme destructiveness. *Journal of Social Issues*, 46(1), 47-64.
- Todorov, T. (1996). *Facing the extreme: Moral life in the concentration camps* (A. Denner & A. Pollak, Trans.). New York: Holt.
- Trivers, R. L. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.
- Webster's desk dictionary of the English language. (1990). New York: Portland House.
- Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.
- Wicklund, R. A. (1975). Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 233-275). New York: Academic Press.
- Wright, D. (1971). *The psychology of moral behavior*. Baltimore: Penguin Books.

Received September 4, 1998

Revision received March 9, 1999

Accepted March 9, 1999 ■