

Walkowitz, Gari; Lönnqvist, Jan-Erik; Irlenbusch, Bernd

**Conference Paper**

## Moral hypocrisy: Self-deception or impression management?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Social Values and Political Attitudes, No. A03-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Walkowitz, Gari; Lönnqvist, Jan-Erik; Irlenbusch, Bernd (2013) : Moral hypocrisy: Self-deception or impression management?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Social Values and Political Attitudes, No. A03-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79701>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

RUNNING HEAD: MORAL HYPOCRISY

Moral hypocrisy: Self-deception or impression management?

Jan-Erik Lönnqvist

Swedish School of Social Science, University of Helsinki

Bernd Irlenbusch

Department of Corporate Development and Business Ethics, University of Cologne

Gari Walkowitz

Department of Corporate Development and Business Ethics, University of Cologne

Keywords: Moral Motivation, Moral Hypocrisy, Impression Management,  
Self-Deception, Personal Values

JEL-Codes: A13, C91, D01, D03, D30, M14

## Abstract

We performed four experiments ( $N = 320$ ) to investigate whether moral hypocrisy is motivated by conscious impression management concerns or whether it is self-deceptive. In a dictator game, moral hypocrisy occurred both within participants (saying one thing, doing another thing; Study 1) and between participants (doing one thing when it is inconsequential, doing another thing when it affects monetary payoffs; Study 2). People were willing to let an ostensibly fair coin decide the monetary allocations only if they could fudge the results of the coin flip, suggesting that moral hypocrites do not deceive themselves (Study 3). Also supporting this view, moral hypocrisy was associated with adherence to Conformity values (Studies 1 and 2), indicative of a desire to appear moral in the eyes of others but not indicative of self-deception. Universalism values were predictive of moral integrity (Studies 1 and 3). Our results help determine how to best reduce moral failures in organizational settings. For example, we show that moral hypocrisy can be reduced by increasing the collective awareness about its prevalence in a specific situation (Study 4).

Moral hypocrisy: Self-deception or impression management?

### Introduction

What motivates moral action, the desire to be moral or the desire to appear moral in the eyes of others? A groundbreaking series of studies conducted by Batson and colleagues uncovered the commonness of moral hypocrisy (MH), the motivation to appear moral yet, if possible, avoid the cost of actually behaving morally (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997; Batson, Thompson, & Chen, 2002; Batson, Thompson, Seufferling, Whitney, & Strongman, 1999). However, failure to behave morally may also carry costs; the awareness that one is acting immorally threatens one's self-concept (e.g., Griffin & Ross, 1991). Guarding against such awareness, and thereby facilitating moral transgressions, is self-deception, a state in which "one's true motivation is masked from oneself" (Batson et al., 1997, *p.* 1346).

In the present study, we sought to contribute to research on the prevalence and mechanisms of MH. By definition, MH involves motivated impression management or other-deception, wanting to give the impression that one is more moral than would be warranted by one's behavior. Furthermore, previous research (Batson et al., 1999) suggests that MH is also self-deceiving, i.e., moral hypocrites also fool themselves. However, we argue that to date, the evidence regarding whether MH is self-deceiving is not decisive. We investigated this issue in a straightforward dictator game setting and found strong evidence of MH both within participants (Study 1) and between participants (Study 2). To distinguish self-deception from other-deception, in Study 3 we altered the experimental setup of Study 1 in such a way that it ruled out disingenuous claims to morality. Study 3 revealed that the proportion of people who were actually willing to commit themselves to behaving morally when they knew that they would have to honor their word was very low, suggesting that moral hypocrites were mostly not self-deceptive.

We also approached the distinction between self-deception and other-deception from the perspective of individual differences. Particularly, the motives underlying MH were investigated within the framework provided by Schwartz' values theory (1992). As we argue below, Conformity values should be indicative of the extent to which one is consciously motivated to maintain a public appearance of being moral. However, self-deceptive processes should be of diminutive importance in the moral decision-making of conformists. Our results showed that adherence to Conformity values was related to both within-participants (Study 1) and between-participants (Study 2) hypocritical behavior, further supporting the importance of motivated other-deception over self-deception. By contrast, Universalism values, indicative of the importance one attaches to fairness and equality, were correlated with more moral behavior (Studies 1 and 3).

As a very practical implication of our findings from Study 1 to 3 in an additional experiment we explored whether making people collectively aware of potential hypocritical behavior helps reducing MH (Study 4). Our results showed that such an intervention indeed diminishes MH and raises the number of fair choices.

Taken together, our results suggest that the impression management motives of moral hypocrites are not complemented by self-deception. In addition to being theoretically important, this result is of high applied interest because it should help determine how to best reduce moral failures in organizational settings; raising self-awareness may not be sufficient if the key problem is a lack of moral motivation. We propose that besides making people collectively aware of potentially hypocritical behaviors, personal values offer organizations a tool with which to promote moral motivation, both in terms helping make staffing decisions, and in terms of being responsive to interventions designed to bring about changes in value priorities.

*Are Moral Hypocrites Self-Deceptive?*

## Moral hypocrisy 5

Why do people sometimes behave morally, even when such behavior appears to be against their self-interests? Primarily, two types of explanations have been offered. Social preference explanations suggest that people have a preference for moral behavior and are willing to pay costs to satisfy this preference (e.g., Fehr & Schmidt 1999). Social signaling explanations, by contrast, suggest that people behave morally to demonstrate to others that they are moral, not because they would actually wish to be moral (e.g., Barclay & Willer, 2007).

Testifying to the importance of the desire to appear moral is a highly influential series of studies conducted by Batson and colleagues (et al., 1997, 1999; 2002). These studies conceptualized the overarching distinction between social preferences and social signaling explanations of morality in terms of moral integrity and MH. Moral integrity was defined as the motivation to actually behave morally. MH, in contrast, was defined as the motivation to appear moral yet, if possible, avoid the cost of actually behaving morally. Moral hypocrites will enact morality not with an eye on producing a good outcome but to appear moral yet still benefit themselves. In a typical design, Batson (et al., 1997, Study 2) had participants assign tasks to themselves and an unknown other participant. One task was described as fun and rewarding, whereas the other was defined as boring. Participants were given the option of flipping a coin to assist in making the decision (this approach was the only way to assign the tasks fairly); however, it was made clear that the coin flip was not required. Typically, approximately half of the participants decided not to use the coin and instead directly assigned the desirable task to themselves. These participants were acting out of self-interest but were not hypocrites because they did not try to conceal their self-interest. Of more interest was the other half of the participants – those who decided to use the coin. Surprisingly, approximately 90% of the coin flippers tossed the better task for themselves. The aggregate data thus clearly indicated that these participants merely claimed to have let the coin decide, suggesting that they were hypocrites.

Batson and colleagues proceeded to ask whether moral hypocrites recognized behaving immorally: did moral hypocrites deceive themselves into believing that they were moral despite their behavior to the contrary? Referring to Alfred Mele's (1987) definition of 'ordinary self-deception' as holding two conflicting views simultaneously, Batson and colleagues argued that people are self-deceived if they can avoid confronting a discrepancy between their behavior and their moral standards. Such 'ordinary self-deception', the authors argued, is sufficient "to reach the goal of appearing moral to one-self" (Batson et al., 1999, *p.* 527). Supporting the idea that moral hypocrisy involves self-deception, Batson et al. (1999) conducted an experiment in which self-awareness, ostensibly working against self-deception, reduced moral hypocrisy. In this study, participants could see themselves in a mirror during the decision task. The introduction of a mirror was argued to increase self-awareness and thereby decrease self-deception. Because the mirror eliminated hypocrisy (in the mirror condition, 5 out of the 10 participants who flipped the coin assigned themselves the positive outcome), Batson et al. (1999) argued that moral self-deception is a prerequisite of moral hypocrisy. Consistent with the idea that self-deception may play a role in explaining moral failures, several other studies (e.g., Shu, Mazar, Gino, Ariely, & Bazerman, 2012; Verplanken & Holland, 2002) have shown that ostensibly raising self-awareness leads to more ethical decision-making. In essence, according to this line of thought, moral failures are facilitated by self-deceptive processes that allow people to maintain their positive view of themselves whilst simultaneously reaping the benefits of transgressing morally.

We believe that the mirror introduced by Batson et al. (1999) may also have had effects other than raising self-awareness. Most pertinent to the distinction between other-deception and self-deception, the presence of a mirror is known to increase the sense of being watched (Fenigstein & Vanable, 1992). More generally, self-attention or self-consciousness causes a heightened feeling of being observed (e.g., Fenigstein & Vanable, 1992) and transparent to

others (Vorauer & Ross, 1999). Importantly, such decreases in the subjective sense of anonymity are known to increase moral behavior (e.g., Zhong, Bohns, & Gino, 2010). This finding means that a decrease in the subjective sense of anonymity, brought about by the introduction of a mirror, could have been responsible for the reduced rates of moral hypocrisy that Batson et al. (1999) reported on.

Based on the above reasoning, we believe that it is necessary to ask whether the conclusion that moral hypocrites engage in self-deceptive processes is warranted. Were those of Batson's participants who flipped the coin actually self-deceptive in the sense that they believed that they would abide by the result of the coin flip whilst simultaneously knowing that they would not? The lucky participants, who let fate do the dirty work for them, and those particularly adept at deceiving themselves (e.g., "mulligan", "best out of three coin flips", see Shalvi, Dana, Handgraaf, & De Dreu, 2011), could eventually persist in this self-righteous belief even after the coin flip. Or were coin flippers choosing to flip the coin merely for the sake of appearing moral, with full knowledge from the beginning that they would fudge the results if necessary? If the latter is true, then the introduction of a binding coin, i.e., a coin flip that cannot be rigged, should dramatically decrease the number of participants choosing to flip the coin.

#### *Personal Values and Moral Behavior*

One way to distinguish self-deception from motivated other-deception is to examine individual difference variables that could be differentially related to these two types of deception (Tetlock & Manstead, 1985). In the present context, this approach means examining how hypocrites differ from those who are more frank about their selfishness: are the characteristics that distinguish moral hypocrites indicative of self-deceptive processes or merely other-deceptive processes? Furthermore, within which framework of individual differences should one search for such characteristics? In a large-scale effort to offer a synthesis of past research in moral



psychology, Graham and colleagues (2011) drew on evolutionary psychology and anthropology to provide a universal taxonomy of the entire moral domain. The authors found empirical support for five universal moral concerns: Harm, Fairness, In-group, Authority, and Purity. The authors noted that four out of these five moral concerns were presented in Schwartz' (1992) values theory.

Values can be defined as transsituational goals that serve as guiding principles in the life of a person or group (Schwartz, 1992). According to Schwartz' model, people in most cultures distinguish between at least ten basic values (the number of cultures in which the same ten basic values have been found was recently reported as 77; Schwartz, 2009) differentiated by motivational content: Universalism, Benevolence, Tradition, Conformity, Security, Power, Achievement, Hedonism, Stimulation, and Self-Direction. Graham et al. (2011) showed that the basic values Benevolence and Universalism were strongly correlated with the universal moral domains of Harm and Fairness, respectively. Similarly strong correlations were found between the basic values Conformity, Tradition, and Security and the moral domains In-group and Authority. Only the moral domain of Purity, related to issues of disgust, naturality, and religiosity, was not represented by any of the ten basic values. From the perspective of the common distinction between individual-level and group-level moral concerns, labeled, e.g., the ethic of Autonomy and the ethic of Community (Shweder, Much, Mahapatra, & Park, 1997), Graham et al. (2011) noted that Harm and Fairness are individual-level moral concerns, whereas In-group and Authority are group-level moral concerns.

Graham et al. (2011) showed that individual-level moral concerns revolve around protecting individuals from harm or unfair treatment by other individuals or the social system. In contrast, group-level moral concerns are focused on creating a tightly knit and ordered community that regulates moral behavior. In terms of this distinction, the motivation to be

considered by others as moral is clearly inspired by group-level moral concerns, which narrows down the list of possibly relevant moral values to Tradition, Security, and Conformity. Of these three, Conformity values should, both on conceptual grounds and based on prior empirical results (reviewed below), be particularly pertinent to the present issue. People high in Conformity values will by definition strive not to upset or harm others and will want to behave according to social expectations or norms (Schwartz, 1992; for empirical evidence, see Lönnqvist, Walkowitz, Wichardt, Lindeman, & Verkasalo, 2009, Study 1). However, their reliance on external guidelines also means that they may have failed to internalize other moral values (Lönnqvist et al., 2009, Study 2; see also Lönnqvist, Leikas, Paunonen, Nissinen, & Verkasalo, 2006). Indeed, whereas certain values, such as Universalism, are thought to give rise to a moral obligation to act morally to preserve one's sense of self-worth (e.g., Beck & Ajzen, 1991; Maio & Olson, 1995; Schwartz, 1977; Schwartz & Tessler, 1972), those people high in Conformity have been shown not to feel morally obligated to behave according to their other moral values (Lönnqvist et al., 2009, Study 3). The self-concept of those adhering to Conformity is thus less sensitive to whether they are acting morally. The individual-level finding that people who value Conformity act less according to their other moral values has recently been replicated at a country level. In a cross-cultural meta-analysis, Boer and Fischer (2013) found that individuals' pro-self vs. pro-social attitudes reflected their corresponding personal values only in more individualistic societies.

Further supporting the view that conformists think of morality in terms of maintaining the conventions provided by society, the level of moral reasoning, as conceptualized either through Kohlberg's (1984) stages or the moral schemes of the Defining Issues Test (Rest, 1979), is inversely related to Conformity values (Helkama, 2004; Myyry et al., 2010). Conformists may feel so strongly obliged by the values and norms that society provides that questions regarding morality seldom if ever arise, as also suggested by some recent results according to which

conformist have difficulties in reporting on a single moral dilemma from their own life (Helkama, 2011).

Consistent with the above view of Conformity values as indicating a motivation to appear moral in the eyes of others, these values clearly show the highest correlations with scores on the Marlowe-Crowne social desirability scale (Schwartz, Verkasalo, Antonovsky, & Sagiv, 1997), a measure of the extent to which the respondent seeks to impress others with his/her good character and gain their approval (Crowne & Marlowe, 1960; see also Paulhus, 1984; 2002). Taken together, the above results suggest that conformists may, if faced with a decision between fair and selfish behavior, lack a genuine preference for fairness but still be highly motivated to appear moral. In light of the above reviewed literature, those adhering to Conformity tend not to look inward for moral guidance, suggesting little, if any, role for self-deceptive processes. It is not themselves that conformists want to deceive but others. In terms of the current design, Conformists would, in the absence of any preference for fairness and with no actual intent to let the coin decide, be expected to claim using the coin.

The two individual-level moral concerns, Harm and Fairness, which are represented in Schwartz' values theory by Benevolence and Universalism, respectively, could be expected to be relevant for actually behaving morally. However, although Benevolence values provide an internalized motivational base for caring for the welfare of others (Schwartz, 1992), this caring is limited to those with whom one is in frequent personal contact; most critical are relations within the family and other primary groups. Because the situations that we investigated were all conducted with anonymous strangers, Benevolence values should not be relevant. In contrast, the defining goal of Universalism values is to care for the welfare of all people (Schwartz, 1992), suggesting that Universalism may be highly relevant in the type of situation currently under investigation.

Those scoring high in Universalism indicate being committed to, e.g., equality, justice, and protection for all people. At face value, these people should, in terms of moral integrity and moral hypocrisy, be motivated by the desire to actually be moral. Consistent with the idea that those scoring high in Universalism may be particularly prone to care about morality, Universalism has been related to (a) moral sensitivity (Myyry & Helkama, 2002); (b) the level of moral reasoning, as conceptualized either by Kohlberg's (1984) stages (Helkama, 2004) or by the moral schemes measured by the Defining Issues Test (Rest, 1979; Myyry, Juujärvi & Pessa, 2010); (c) emotions that motivate moral behavior, such as empathy and guilt (Silfver, Helkama, Lönnqvist, & Verkasalo, 2008); and (d) fairer behavior in standard experimental economics games (Lönnqvist, et al., 2009; Lönnqvist, Walkowitz, Verkasalo, & Wichardt, 2011). Based on these findings, those high in Universalism were more often expected to choose the fairest course of action (in our design, a distributively fair allocation of money; see below).

#### *Purpose of the Present Research*

We conducted four studies using a straightforward dictator game to augment our understanding of the mechanisms behind the type of moral failures that Batson and colleagues (et al., 1997, 1999; 2002) have reported on. The dictator game resembles the situation that Batson's participants faced: the dictator decides on how to split a given amount of money, and the other party is passive; we needed a passive person to ensure the credibility of our decision task in the eyes of our participants.

In Study 1, we employed a within-participants design similar to those designs used by Batson and colleagues (et al., 1997; et al., 1999; et al., 2002) to examine moral failure revealed by inconsistency between what one says and what one does. However, to more clearly discern between the motivation to be moral and to appear moral – the two motivational goals of Universalism and Conformity values, respectively – we introduced a distributively fair option into the design: Dictators could either directly choose a distributively fair (5/5) or selfish (8/2)

allocation of money, or they could let a coin flip decide the allocation. In contrast to Batson's studies, those who genuinely preferred to be fair could directly choose the distributively fair allocation. We denote choosing to flip the coin as seemingly procedurally fair; although the procedure could at face value be considered procedurally fair, the results later revealed that participants did not generally abide by the result of the coin flip.<sup>1</sup> Based on the results of Batson's research program, we expected a substantial number of participants to claim using the coin but to reveal, through their aggregate behavior, that at least some of them did not abide by the result of the coin flip. More pertinent to the present research, among participants who chose the selfish allocation, we expected those high in Conformity to claim to have flipped the coin to arrive at the selfish outcome. We also expected participants high in Universalism values to be more likely to directly choose the distributively fair allocation of money.

Perhaps the most common way to examine the interpersonal vs. intrapsychic nature of a process is to vary the privacy of the behavior. To investigate the effects of privacy, we included a manipulation in Study 1 designed to heighten the dictators' sense of being watched. We expected choosing the coin to cover up selfish motives to increase in the condition of being watched. Furthermore, we expected conformists to be particularly sensitive to this manipulation (Keller & Pfattheicher, 2011; Leikas, Lönnqvist, Verkasalo, & Lindeman, 2009).

Study 2 was conducted to further affirm that conformists are primarily motivated by the goal to appear moral, not be moral. We now employed a between-participants design to investigate MH (see Monin & Merrit, 2011). Specifically, we expected a between-participants manipulation of monetary incentives to reveal dishonest claims to morality: half of the participants made a hypothetical decision, and half of the participants made a monetarily incentivized decision. Those participants who were asked hypothetically how they would split the money were expected to declare distributively fairer allocations than those participants who

actually allocated the money (the option to let the coin decide was not given in Study 2). More importantly, we expected conformists to be particularly likely to adjust their behavior as a function of whether real money was at stake. Such a result would testify to the importance of impression management motives in making dishonest claims to morality. We also expected, across conditions, those participants high in Universalism to be more likely to select the distributively fair option.

For more decisive evidence regarding the role of impression management and self-deception in moral hypocrisy, we introduced the binding coin in Study 3. Our purpose was to investigate whether Study 1 participants were honestly intending to let the coin decide (self-deception) or whether they were flipping the coin with the intent to falsify the outcome if necessary (impression management). The research design was otherwise similar to the one employed in Study 1, with the exception that participants could not rig the result of the coin flip. Would participants choose to flip a coin that could not be rigged, thereby indicating an honest, although self-deceptive, motivation to behave morally? Based on our above-argued view that moral hypocrites are not necessarily self-deceptive, we expected less participants in Study 3 to flip the coin. We again expected Universalism values to predict choosing the distributively fair option at the outset.

To further investigate the role of impression management in morally hypocritical behavior from a more applied perspective, in Study 4 we again adapted our experimental setup from the watched condition of Study 1 by providing all participants (decision makers and passive persons) with information about the behavior of coin users in the watched condition of Study 1. Our manipulation was intended to make participants collectively aware of the hypocritical coin use in Study 1, i.e., we wanted to make sure that they knew that others knew that the coin was mainly used as an excuse for implementing the unfair outcome. We expected that by our

manipulation such kind of impression management would be perceived as losing its force in the eyes of others and therefore would occur less often.

To summarize, our hypotheses are as follows:

H1 (Study 1). Based on the results reported by Batson and colleagues (et al., 1997; et al., 1999; et al., 2002), we expect most dictators to end up making the selfish allocation (H1a). Of these dictators, approximately half are expected to directly choose the selfish option, whereas approximately half are expected to choose the seemingly procedurally fair option, the coin flip (H1b). The aggregate behavior of the latter group is expected to reveal that these dictators did not abide by the result of the coin flip (H1c).

H2 (Study 1). Among those participants who behave selfishly (end up with the 8/2 allocation), particularly those high in Conformity will claim to have flipped the coin to arrive at the selfish outcome.

H3 (Study 1). Conformists will be more likely to claim to have flipped the coin in the public, compared to the private, condition.

H4 (Study 2). Participants making hypothetical allocations of money will be fairer than those making real allocations.

H5 (Study 2). Conformists will be particularly likely to adjust their behavior as a function of whether hypothetical or real money is at stake.

H6 (Studies 1 and 3). If the result of the coin flip is made binding, participants will no longer select to flip the coin.

H7 (Studies 1 and 4). Making participants collectively aware of the hypocritical use of the coin reduces its force as a means of impression management. Thus, the use of the coin as an excuse for implementing the unfair outcome will be reduced.

H8 (Studies 1 to 4). Universalism values will predict directly choosing the distributively fair allocation of money.

### Study 1

Study 1 represents a modification of Batson's research design in the form of a dictator game. Our purpose was to connect moral hypocrisy – as revealed through within-participants behavioral inconsistency – and moral integrity to Conformity and Universalism values, the former being indicative of impression management motives, the latter of moral integrity.

#### *Methods*

*Participants and Procedure.* Study 1 was conducted with 128 participants (mean age = 24.3 years ( $SD=2.87$ ); 52.3% female) from the University of Cologne (Germany) majoring in different disciplines. Participants were invited via a Laboratory of Economic Research mailing list with approximately 3700 subscribers who had signed up to take part in experiments organized at the laboratory. Four experimental sessions were run involving 64 participants per condition and 32 participants per session. Upon arrival, participants were randomly seated in computer cubicles that secured anonymity, randomly assigned a role (dictator or passive other participant), and randomly matched two participants in dyads. All experimental sessions were conducted on the computer using the experimental platform z-Tree (Fischbacher, 2007). The experiment lasted somewhat less than an hour (participants completed some filler items and the personal values questionnaire after the dictator game). Subsequently, participants were compensated with a fixed amount of 2.5€ (for showing up and completing the questionnaires) along with the amount that they earned in the dictator game.

*Measures.* In the dictator game, two participants were randomly matched and assigned either the role of the dictator or the passive other person. At the outset of the game, the dictator was provided with an amount of 10€ that was to be distributed among herself and the other person. All dictators were additionally provided with a coin that was labeled “8/2” on one side



and “5/5” on the other side (Batson et al., 2002, Study 2). The dictator could choose from three options: 1) she could either directly choose the option “8/2” that yielded 8€ for herself and 2€ for the other person; 2) she could directly choose the option “5/5” that yielded equal payoffs of 5€ for both; or 3) she could flip the coin and determine the distribution according to the outcome of the coin flip. The dictator was asked to indicate her choice (“8/2” or “5/5”) and whether she had flipped the coin (“Yes” or “No”). The other person did not make any decisions. We chose the distribution 8/2 as our selfish option because the mean transfer in dictator games is approximately 20 percent of the endowment (Camerer, 2003), suggesting that such a division is one that many participants would feel comfortable with.

Furthermore, we included a manipulation designed to induce the feeling of being watched. Before acting, half of the dictators were informed that the other person would, via the computer network, be notified of whether the dictator indicated having used the coin (watched condition). In the other condition, it was clear to the dictators that the other person was not notified about the coin use (unwatched condition).

Our decision to adapt Batson’s design rather than directly replicate it was not solely motivated by our wish to include a distributively fair choice. This choice, as argued above, should enable us to more clearly discern between the motivation to be moral and to appear moral. In addition, we took care to address some recent criticism suggesting that Batson’s results on moral hypocrisy are, in fact, a product of experimenter demand effects: Particularly, making the norms of justified self-interest, morality, and fairness salient in the instructions of the experiment has been shown to cause experimenter demand effects that could account for Batson’s findings (Fernández-Dols et al., 2010). For instance, the phrase “[t]he decision is entirely up to you, you can assign yourself and the other participant however you choose” was shown to justify private self-interested behavior and cause the private egoism observed in

Batson's design (Fernández-Dols et al., 2010). We tried to avoid any such framing effects by using as neutral language as possible to describe the task (e.g., by avoiding terms such as "fair"; see the Appendix for full instructions).

We measured personal values with the Portraits Values Questionnaire (PVQ; Schwartz et al., 2001). This measure was, across studies, administered after various filler items at the end of each session. This 40-item questionnaire designed to measure the ten basic values and consists of descriptions of a hypothetical person in terms of his or her goals and aspirations. Respondents are to rate their similarity to the person on each item, using a 6-point scale ranging from 1 (*not like me at all*) to 6 (*very much like me*). The measure includes four items for measuring Conformity (example item: It is important to her always to behave properly. She wants to avoid doing anything people would say is wrong) and six items for measuring Universalism (example item: She wants everyone to be treated justly, even people she doesn't know. It is important to her to protect the weak in society). To compute the values scores, we used proportional sum variables that centered, within participants, the average of the ten basic values at one (Schwartz, 1992). The descriptive statistics for Conformity were  $M = 0.89$ ,  $SD = 0.18$ ,  $\alpha = .64$  and for Universalism  $M = 1.04$ ,  $SD = 0.18$ ,  $\alpha = .77$ . The alpha value .64 could be considered somewhat low. However, as argued by Schwartz et al. (2001) in the validation of the PVQ, low internal consistency reliabilities are to be expected because the basic personal values, although each measured with only a couple of items, have conceptually broad definitions, encompassing multiple components. All alphas we report in the present paper exceed those reported in the validation of the PVQ (in the validation paper, Conformity and Universalism showed reliabilities of .48 and .57, respectively; Schwartz et al., 2001). Furthermore, low internal consistency reliabilities should not generally be interpreted as threatening measurement validity: McCrae, Kurtz, Yamagata and Terracciano (2011) argued and showed that internal consistency reliability was not related to various validity criteria.

### Results

Sex, age, and condition had no effects on any of the dependent variables and were collapsed for correlational analyses. However, all statistically significant correlations were further investigated in regression analyses that controlled for both the condition (watched vs. unwatched) and the interaction between the condition and the investigated personal value (condition x value interaction), as well as for age and sex. As expected (H1a), out of the 64 participants who were assigned the role of the dictator, 56 chose the selfish option (88%). Also as expected (H1b), approximately half – 26 (46%) – of those who chose the selfish option indicated having flipped the coin (behavioral decisions across the three studies are shown in Table 1). Out of the 26 participants who indicated having flipped the coin, all reported having flipped the selfish outcome (8/2). The likelihood that a fairly flipped coin would give such an outcome is extremely small ( $p < .0000001$ ), suggesting, as expected (H1c), that participants were not being truthful.

As expected (H8), Universalism ( $r(63) = .29, p = .02$ ) was correlated with choosing the distributively fair option at the outset (choosing 5/5 was coded as 1; choosing the selfish outcome, with or without coin flip, was coded as 0). Conformity values were not correlated with choosing the distributively fair option ( $r(63) = .06, p > .10$ ). In a regression analysis that predicted choosing the distributively fair option and with condition (watched vs. unwatched), age, sex, Universalism, and the interaction between condition and Universalism values entered as predictor variables, the main effect of Universalism values was statistically significant (Beta = .45,  $t(58) = 2.48, p = .02$ ). No other variables emerged as statistically significant predictors (see Table 2).

Among the 56 participants who chose the selfish outcome, Conformity values were, as expected (H2), correlated with claiming to have used the coin to obtain this outcome ( $r(55) =$

.27,  $p = .05$ ). Conformity remained a statistically significant predictor in a regression analysis that included, in addition to Conformity (Beta = 0.42,  $t(50) = 2.18$ ,  $p = .03$ ), condition (Beta = -1.12,  $p > .10$ ), age (Beta = -.01,  $p > .10$ ), sex (Beta = .14,  $p > .10$ ), and the interaction between condition and Conformity (Beta = 1.07,  $p = .14$ ); this interaction term, contrary to our prediction (H3), was not statistically significant.

### *Discussion*

The results of Study 1 provide firm evidence of moral failure as revealed through behavioral inconsistency. As expected, most dictators ended up making the selfish decision (H1a). Similar to the results of Batson and colleagues (et al., 1997; et al., 1999; et al., 2002), and also as expected (H1b), approximately half of the participants indicated having flipped the coin to determine the distribution of money. The similarity of our results with those reported by Batson and colleagues (et al., 1997; et al., 1999; et al., 2002) is all the more striking given that we, in contrast to previous studies, gave our participants the chance to directly choose the distributively fair outcome. Also as expected (H1c), all 26 coin flippers ended up with the selfish outcome. Among those participants who chose the selfish option, those high in Conformity were, as expected (H2), more likely to use the coin flip as an excuse for their selfish behavior. Considering previous work on Conformity (see Introduction), this result suggests that others, not the self, were the primary target of the deception. Thus, impression management concerns determined the behavior of coin flippers, with little role for self-deception. In contrast, Universalism values were correlated with choosing the distributively fair option from the outset, suggesting that this value can help explain moral integrity. Contrary to our expectations (H3), the watched vs. unwatched manipulation had virtually no effects, suggesting it may have been too weak to induce a feeling of being watched.

The most important difference to earlier studies on moral hypocrisy was that we offered a distributively fair (5/5) option from the outset. Introducing this option allowed us to disentangle

those participants who wanted to be moral (those motivated by Universalism values) from those who merely cared about appearing moral (those motivated by Conformity values). By comparison, in the design employed by Batson there was no distributively fair option; the equivalent in terms of the present design could have been to force participants to choose between 8/2 and 2/8 rather than between 8/2 and 5/5. We decided against including the 2/8 option because we reasoned no one would choose this option; dictators never tend to allocate more than half of the amount to the other person (Camerer, 2003). Because we offered a distributively fair option from the outset, choosing the coin was already the less moral course of action. In the presence of a distributively fair option, the excuse provided by a procedurally fair option has previously been suggested not to be a sufficiently compelling excuse to appeal to those people seeking to cover up their selfish motivation (Batson et al., 2003). However, similar to Batson's results, almost half of our participants did try to hide their selfish motive behind the coin flip. Perhaps, outweighing the ostensible feebleness of the excuse was the steeper threshold of directly choosing the selfish option: self-serving decisions are more difficult to make in the presence of a fair option (Scheepers, Spears, Manstead, & Doosje, 2009). The introduction of an outcome-fair option could thus have had two effects that worked in opposite directions: although such an option makes the excuse offered by the process-fair option (coin flip) look feeble, it also makes it more difficult to directly choose the selfish option. Interestingly, participants did not generally react to the potential unattractiveness of these two options by choosing the third, distributively fair option (only 13% of Study 1 participants chose the third option). Instead, testifying to the pervasiveness of lack of moral integrity, nearly all participants continued to choose the selfish outcome and approximately half of them justified their behavior with the coin. When considering the generalizability from the laboratory context, the possibility to make a distributively fair split may arguably serve to increase the external validity of our research: in many real-life conflicts of

interest, an outcome that is mutually perceived as fair is often implementable. Raising external validity was another reason why we wanted to give participants a distributively fair option from the outset.

As noted above, Fernández-Dols and colleagues (2010) provided evidence that the hypocrisy revealed by Batson's research could be a phenomenon produced by experimenter demand effects. The authors suggested that in real life, hypocrisy, because of its social dysfunctionality and the derogation of hypocrites, is infrequent. Therefore, although our experimental design was different than Batson's, a comparison between the numbers of participants who displayed moral hypocrisy may still be interesting. In Study 1, 40% ( $N = 26$ ) of our participants flipped the coin, and many of these participants were with high probability hypocrites, as revealed by the fact that 100% of these 26 participants flipped the selfish option for themselves. These numbers are very similar to the ones reported by Batson and colleagues. In the setup closest to our research design, Batson et al. (1997; study 2) reported that 10 out of 20 (50%) participants flipped the coin, and nine of those ten (90%) assigned themselves the fun task. Thus, although our research design did not involve those demand effects that Fernández-Dols and colleagues (2010) identified as possibly causing the moral hypocrisy phenomenon, our design pointed to the high frequency of moral hypocrisy.

## Study 2

Study 1 investigated within-participants moral hypocrisy (the same people claiming one thing but doing another). In Study 2, we used a between-participants manipulation of monetary incentives to expose moral hypocrisy (doing one thing when it is inconsequential and another thing when it affects monetary payoffs; see Monin & Merrit, 2011) and expected those high in Conformity, primarily motivated by the goal of appearing moral, to be particularly sensitive to this manipulation.

### *Methods*

*Participants and Procedure.* Study 2 was performed with 64 participants (two sessions with 32 participants; average age=24.2,  $SD=2.61$ ; 51.6% female) from the University of Cologne. The procedure was similar to that of Study 1.

*Measures.* The dictator game in Study 2 was similar to the one in Study 1, except that there was no coin and both players made a choice as dictators. After receiving an endowment of 10€, the incentivized dictator made a monetarily incentivized decision between the “8/2” or the “5/5” distribution. The hypothetical dictator made the corresponding hypothetical choice (hypothetical condition).

Personal values were measured by the Portraits Values Questionnaire (Schwartz et al., 2001). The descriptive statistics were for Conformity  $M = 0.89$ ,  $SD = 0.21$ ,  $\alpha = .74$  and for Universalism  $M = 1.08$ ,  $SD = 0.23$ ,  $\alpha = .84$ .

### *Results*

Age and sex had no effects on any of the outcomes. Regarding the effect of condition, out of the 32 participants who made an incentivized choice, 26 chose the option that favored themselves, whereas only six chose the distributively fair option. By contrast, in the hypothetical condition, slightly more than half of the participants, 17 out of 32, chose the distributively fair option. As expected (H4), the distributively fair option was chosen more often in the hypothetical than in the incentivized condition ( $\chi^2(1) = 6.78$ ,  $p = .01$ ).

To examine the effects of Conformity, we ran a regression analysis predicting choice behavior with condition (incentivized vs. hypothetical), age, sex, Conformity, and the interaction between Conformity and condition (see Table 2). The expected interaction (H5) between Conformity and condition was statistically significant (OLS regression: Beta = 0.99,  $t(58) = 2.00$ ,  $p = .05$ ). The simple slopes from the regression are plotted in Figure 1. As expected, in the

hypothetical condition, with age and sex also entered as predictor variables, Conformity values predicted choosing the distributively fair option (Beta = 0.32,  $t(28) = 2.01$ ,  $p = .05$ ). However, Conformity did not predict this choice in the incentivized condition (Beta = -.08,  $t(28) = -.42$ ,  $p > .10$ ). A similar regression analysis, with Universalism replacing Conformity as independent variable, did not reveal the expected effect (H8) of Universalism (see Table 2).

### *Discussion*

The results of Study 2 showed, as expected (H4), that many participants were motivated to convey, without cost, an appearance of being moral. However, in the presence of real stakes, very few were willing to pay the price for actually behaving morally. In essence, slightly more than half of the participants in the hypothetical condition claimed that they would choose the distributively fair rather than the selfish option. However, the disingenuous nature of this claim (i.e., participants were distorting their responses to appear moral) was revealed in the monetarily incentivized condition, in which participants were more selfish.

As expected (H5) and consistent with the results of Study 1, particularly those participants scoring high in Conformity values made disingenuous claims to morality. The lack of findings for Universalism is most likely explained by the very small amount of players who chose the 5/5 distribution.

### Study 3

Studies 1 and 2 showed that Conformity values, with the desire to appear moral at their core, were associated with MH. These studies thus suggested that MH may be more a matter of motivated impression management than self-deception; conformists are unlikely to experience a need to deceive themselves. To further establish the lack of self-deception among those participants who flipped the coin in Study 1, we introduced the binding coin in Study 3. Would participants select to flip the coin if the result of the coin flip could not be fudged? If not, this



outcome would suggest that those participants who flipped the coin in Study 1 were simply engaging in impression management.

### *Methods*

*Participants and Procedure.* Study 3 was executed with 64 participants (32 participants per session; average age=24.5,  $SD=3.1$ ; 56.3% female) from the University of Cologne. The procedure was similar to that of Study 1.

*Measures.* We changed our experimental setup from Study 1 by enforcing the coin flip. In essence, participants who chose the coin flip to determine the outcome were forced to abide by the result of the coin flip. At the outset of the game, the dictator again received an endowment of 10€ that was to be distributed between herself and the passive other person. As in Study 1, the dictator could 1) directly choose the option “8/2”, 2) directly choose “5/5”, or 3) shift her decision to a fair lottery implemented by the experimenters. This lottery (which was run on the participant’s computer) was structurally identical to the coin in Study 1 in that it yielded with a probability of 50% the option “8/2” and with a probability of 50% the option “5/5”. The other player did not take any action.

Personal values were measured with the PVQ (Schwartz et al., 2001). The descriptive statistics were for Conformity  $M = 0.88$ ,  $SD = 0.20$ ,  $\alpha = .53$  and for Universalism  $M = 1.03$ ,  $SD = 0.15$ ,  $\alpha = .75$ .

### *Results*

Age and sex had virtually no effects on any of the outcomes. Out of the 32 decision-makers, six chose the distributively fair option at the outset, 22 chose the selfish option, and four (12.5%) chose the binding coin flip. Compared to Study 1, in which 26 participants out of 64 (40.6%) flipped the coin, the number of participants who now chose the coin was, as expected (H6), significantly smaller (comparing across studies,  $\chi^2(1) = 6.60$ ,  $p = .01$ ).

Universalism values predicted choosing the distributively fair option at the outset ( $r(31) = .37, p = .04$ ; directly choosing the distributively fair option was coded as 1, other choices as 0). Universalism retained its predictive power in a regression that also included age and sex as predictor variables (Beta = .41,  $t(28) = 2.29, p = .03$ ; see Table 2).

### *Discussion*

Comparing the results of Study 3 with those of Study 1 shows, consistent with H6, that although half of the participants claimed to have abided by the result of a coin flip (Study 1), very few were actually willing to commit themselves to a binding coin (Study 3). This result means that the decision in Study 1 to flip the coin was most likely determined by the motivation to appear moral. Participants in Study 1 did apparently not deceive themselves into believing that they would actually be moral and let the coin decide. Rather, these participants were prepared to choose the coin only if the opportunity to fudge the results was left open.

Regarding personal values, the number of participants who chose the binding coin flip was rather small, not allowing for conclusions regarding the values of those people who made this choice. However, consistent with the results of Study 1, participants high in Universalism were, as expected (H8), more likely to choose the distributively fair option at the outset.

### Study 4

Our studies 1 to 3 have shown that MH is more a matter of motivated impression management and not primarily of self-deception. We also demonstrated that people high in Conformism values are more prone to such other deceiving behavior.

To assess a practical implication from these insights, we investigated whether making people collectively aware of the hypocritical use of the coin would reduce MH. We hypothesized that if decision makers (dictators) as well as those affected by their decisions (passive persons) knew about the frequent hypocritical use of the coin as an excuse to implement the unfair outcome, such impression management would be less effective and therefore would be exhibited

less often, i.e., decision makers would less often claim to have used the coin and/or coin users would be more likely to implement 5/5. To investigate this, in Study 4 we provided all participants with information about coin users' decisions from the watched condition of Study 1 before they engage in the same decision task as the participants in the watched condition of Study 1.

### *Methods*

*Participants and Procedure.* Study 4 was executed with 64 participants (32 participants per session; average age=24.8,  $SD=3.1$ ; 68.8% female) from the University of Cologne. The procedure was similar as in Study 1.

*Measures.* We marginally modified our experimental setup from the watched condition in Study 1 by providing participants with information about the behavior of coin users in the watched condition of Study 1. We conveyed that all coin users reached the unfair outcome and no coin users implemented the fair outcome. We also made sure that all participants knew that all others participants had received this information. The subsequent distribution task was the same as in Study 1. Personal values were measured with the PVQ (Schwartz et al., 2001). The descriptive statistics were for Conformity  $M = 0.92$ ,  $SD = 0.20$ ,  $\alpha = .76$ , and for Universalism  $M = 1.07$ ,  $SD = 0.19$ ,  $\alpha = .67$ .

### *Results*

Age and sex had no effects on any of the outcomes. Out of the 32 dictators, 7 (21.8%) chose the distributively fair option at the outset, 14 (43.8%) chose the selfish option, and 11 (34.4%) chose to flip the coin.

Moral hypocrisy decreased in Study 4: Firstly, 21 participants did not use the coin and 4 participants used the coin and produced the fair outcome. This amounted to a total of 25 participants who clearly avoided being regarded as hypocrites. This number was significantly

higher than the 16 participants in Study 1 who could not be suspected of being a hypocrite ( $\chi^2(1) = 5.50, p = .02$ ). Secondly, focusing on the coin users we found that the distributions of outcomes differ significantly between Study 1 and Study 4. In Study 1 all 16 coin users reached the unfair outcome while 4 coin users implemented the fair outcome in Study 4 ( $\chi^2(1) = 6.83, p < .01$ ) suggesting that in Study 4 coin users forge the coin results considerably less often. As a consequence significantly more participants implemented the fair outcome in Study 4 (11; 34.4%) than in Study 1 (4; 12.5%) ( $\chi^2(1) = 4.27, p = .04$ ).

The predicted (H8) correlation between Universalism values and choosing the distributively fair option at the outset did not reach statistical significance ( $r(31) = .23, p > .10$ ; directly choosing the distributively fair option was coded as 1, other choices as 0). Neither were Conformity values correlated with choosing the distributively fair option ( $r(31) = -.16, p > .10$ ). However, in a regression analysis that predicted choosing the distributively fair option with age, sex, and Universalism values, the predicted effect of Universalism values was marginally statistically significant (Beta = .34,  $t(28) = 1.81, p = .08$ ). No other variables emerged as statistically significant predictors (see Table 2). A similar regression analysis, with Conformity replacing Universalism as independent variable, revealed no effects (see Table 2).

### *Discussion*

As expected (H7) our results suggest that social awareness of the hypocritical use of the coin in Study 1 weakened the force of impression management that builds on using the coin as an excuse to implement the unfair outcome. Interestingly, our intervention in Study 4 had similar effects as the mirror intervention conducted by Batson et al. (1999); i.e., more participants decided to not use the coin and fewer of those implemented the selfish outcome. Additionally, fewer coin users implemented the selfish outcome.

The results from Study 4 are in line with our main finding of the other three studies that impression management is the decisive underlying mechanism of hypocritical behavior. A

potentially important practical implication emerges: making the frequent hypocritical use of seemingly fair procedures publicly known appears to be a very effective and easy to implement measure to reduce MH and as a consequence to enhance fair behaviors.

### General Discussion

Our results showed that many people would not want to pay the cost of behaving morally if they can avoid doing so without appearing immoral. Particularly, in Study 1, almost half of the participants claimed to have flipped the provided coin to assign money. However, undermining their credibility, everyone claimed to have flipped the self-favoring outcome. In Study 2, people were prone to split hypothetical money equally. However, in the monetarily incentivized condition, people generally favored themselves. In Study 3, when faced with a binding coin flip, participants were generally not willing to commit themselves to the coin but preferred directly choosing the selfish option. People were thus willing to flip the coin if they knew from the beginning that the results could be fudged (Study 1) but not when this was impossible (Study 3). This finding suggests that participants were not self-deceptive when choosing to flip the Study 1 coin; participants did not believe that they would abide by the result of the coin flip

Also suggesting the importance of motivated impression management concerns over self-deception, particularly those participants scoring high in Conformity values were prone to behave hypocritically. It was those participants who claimed to have flipped the coin to assign the money (Study 1) and, in the absence of monetary costs, that they would choose a distributively fair outcome over a self-favoring one (Study 2). The dishonesty of this latter claim was revealed in the monetarily incentivized condition, in which conformists were not more likely to behave morally. Publicly revealing the use of the coin as an excuse to implement the selfish outcome reduced the frequency of this hypocritical behavior (Study 4). Combined, the results of these four studies suggest that hypocrites are aware of their moral failures; they are primarily

motivated by interpersonal, not intrapsychic, concerns.

*The Riddle of Self-Deception*

The riddle of self-deception – can one simultaneously believe both  $p$  and  $\neg p$ ? – holds an enduring appeal for philosophers and psychologists. Our contribution to this discussion was to investigate the more specific assertion that self-deception is a prerequisite of moral hypocrisy (Batson et al., 1999; see also von Hippel & Trivers, 2011). First, our studies revealed a fair amount of other-deception. In Studies 1 and 2, participants were being dishonest about how they had arrived at the selfish decision (Study 1) and about how they would behave if there were money at stake (Study 2). However, such deception, it appears, was not accompanied by self-deception: When people chose the coin flip, they were relying on their ability to fudge the results if necessary (Study 3). Also supporting this interpretation, particularly those participants scoring high in Conformity values, i.e., those who, when faced with an ethical decision, would not care whether they are moral but who would nevertheless want to appear moral in the eyes of others, were prone to behave hypocritically (Studies 1 and 2). It appears, therefore, that the observed hypocrisy was primarily inspired by a conscious attempt to impress the anonymous other participant or an unknown experimenter and not by a self-deceptive process aimed at sustaining one's self-image as a moral person.

Although variants of the proposal that other-deception is accompanied by self-deception have been put forward for quite some time (e.g., Trivers, 1985, 1991), there is very little relevant empirical data on this hypothesis, as also acknowledged in a recent review (von Hippel & Trivers, 2011). One reason for this lack of empirical research could be the general difficulty of discriminating between impression management and intrapsychic explanations. As argued by Tetlock and Manstead (1985), some deep-rooted theoretical and conceptual issues may always render this distinction more or less arbitrary. For instance, in the face of empirical data that suggests little or no external influence on the investigated behavior, intransigent impression

management explanations can posit that people try to impress internal audiences. Similarly, external manipulations of the situation can always be argued to affect not only impression management considerations but also intrapsychic processes. Underlying the distinction between impression management and self-deceptive explanations of behavior is the more general tension between outer and inner explanations of behavior. This tension runs through the history of personality and social psychology and appears unlikely to be settled by empirical data, which may explain the reluctance of empirically minded researchers to investigate proposals regarding the prevalence and functions of self-deception.

Related to the above-referred difficulty of disentangling impression management from intrapsychic explanations, manipulations designed to affect only one type of process but not the other are challenging to conceive of. As noted in the Introduction, a heightened self-awareness is generally accompanied by a heightened feeling of being watched and being transparent to others (Fenigstein & Vanable, 1992; Vorauer & Ross, 1999). Not only mirrors but also other typical manipulations designed to raise self-awareness, such as signing (Charness & Gneezy, 2008), may increase the sense of being watched. In fact, one of the general consequences of increased self-awareness may be a decrease in the subjective sense of anonymity (e.g., Festinger, Pepitone, & Newcomb, 1952; Zimbardo, 1969). If an increase in self-awareness is associated with a decrease in subjective anonymity, then it could be that the latter is doing the causal work when self-awareness is manipulated; there is firm evidence that anonymity is related to moral transgressions (e.g., Zhong, Bohns, & Gino, 2010).

In part because of the above-described challenges in specifying which of the effects of self-awareness manipulations are causally relevant, we chose another approach to the disentanglement of self-deceptive and impression management motives. Focusing on individual differences in Conformity values – which we argued would be indicative of a tendency to care

about the impression one makes on others, not of an inclination to deceive oneself – we showed that the impression management of moral hypocrites may generally not be accompanied by self-deception. The more general point we wish to make is that individual difference variables may sometimes be useful for examining otherwise elusive cognitive processes. Although this approach, just as any other, suffers from limitations, it is reassuring if different methods produce similar results. Individual differences research should, therefore, more often be used to complement experimental research, to the mutual benefit of both research traditions.

Our reliance on individual differences to examine a cognitive process does recommend some caution in interpreting our results. Our results do not suggest that self-deception never accompanies other-deception. Rather, our results suggest that under these particular conditions, involving an objectively observable moral behavior, impression management processes are of primary importance and intrapsychic processes (self-deception) are less important. In essence, people are concerned about the social identity implications of their decisions, not the personal appropriateness or correctness of their decisions. This finding resulted despite the fact that other aspects of the research design, such as anonymity and the one-shot nature of the interaction, could have been expected to weight against impression management concerns.

#### *Contribution to other Areas of Behavioral Ethics Research*

Our results could, at first glance, be interpreted as being at odds with some previous results suggesting that raising self-awareness will cause more moral behavior. In fact, our data are not inconsistent with these previous results. However, our data do suggest that the mechanism by which increased self-awareness functions to make ethical standards more salient may not involve a decrease in self-deception. First, as noted above, an increase in self-awareness is typically accompanied by a decrease in anonymity, which could induce more moral behavior. Second, personal values influence behavior only when activated (e.g., Maio, 2010). One way to activate values is by increasing self-awareness (e.g., Verplanken & Holland, 2002). Thus,



another mechanism through which increased self-awareness could lead to more moral behavior could be that self-awareness makes personal values more salient. Whether either of these two mechanisms or both are involved in the process that leads from self-awareness to more moral behavior should be an intriguing question for future research.

Our results could also be interpreted as being at odds with the literature on moral awareness; there is firm empirical evidence that making people aware of the moral dimensions of their decisions leads to more moral decisions (e.g., Bazerman & Gino, 2012). In fact, we agree with this literature; however, we wish to add the caveat that raising moral awareness might only lead to more moral action when all involved people share this awareness and when all people can assume that all others share this awareness. Moreover, at the individual level, raising moral awareness might only enhance moral action in some people, particularly, in those people who value the well-being of others. Previous research on the influence of values on behavior suggests that making certain values more salient will only influence behavior if those values are important to the person (for a review, see Maio, 2010). This finding means that raising the moral awareness of those people high in Conformity would most likely not affect their behavior. By contrast, raising the moral awareness of those people who care for the well-being of others should lead to increases in moral behavior.

Our results are consistent with recent research on moral balancing. A considerable amount of research on moral balancing or moral licensing suggests that prior good deeds cause people to feel licensed to subsequently act immorally (e.g., Merritt, Effron, & Monin, 2010). However, some recent results also suggest an important boundary condition to this general rule; this rule only holds true if the prior good deeds are publicly recognized (Sussewind, Hoelzl, & Walkowitz, 2012). This result is consistent with the results of the present research; if people primarily care about being perceived as good, as many people do according to the present results,

then only publicly recognized good deeds should free people to be bad.

On a general level, our results testify to the strength of social signaling explanations of moral behavior. It is important to note that if self-deception were a prerequisite for MH, as Batson et al., (1999) claimed, then hypocritical behavior would, although regularly cited as a prime example of the importance of social signaling, be possible only if it also satisfied the motive to perceive oneself as fair. However, in light of our results, MH does not involve self-deception and, therefore, does not serve social preferences. Our results also suggest that both social preferences and social signaling can explain moral behavior; some people may be more motivated by the former (those adhering to Universalism) and others more by the latter (those adhering to Conformity).

#### *Applied Relevance for Organizational Behavior*

In addition to advancing our theoretical understanding of a very subtle and complex issue that has attracted the attention of philosophers for centuries – why do ostensibly good people behave immorally – our results could also have applied relevance. Whether impression management motives are accompanied by self-deception may be important in determining how to best help reduce moral failures. Such failures have recently become quite commonplace across a variety of industries and occupations, leading to an increased interest in business ethics training both for professionals and business school students. However, a recent review and meta-analysis on business ethics instructional programs suggests that they have at best a minimal impact (Waples, Antes, Murphy, Connelly, & Mumford, 2009). Our results offer some suggestions for the development and improvement of these programs.

If self-deceptive processes play an important role in the moral life of hypocrites, then focusing on raising self-awareness could be an effective means of countering immoral behavior. However, the absence of such self-deceptive processes would suggest that hypocrites are already aware of their wrongdoings and merely lack the motivation to do what is right. In light of our

results, fostering the motivation to do what is morally right and publicly discussing the potentially immoral nature of seemingly moral behaviors should be amongst the top priorities of business ethics courses. Personal values constitute strong motivational forces that due to their abstract nature are thought to guide behavior across a variety of contexts and situations. Focusing on value change could therefore be an efficient means by which to strengthen the motivation to behave morally.

Prior research on personal values has typically built on the assumption that values are relatively stable. However, a recent review by Bardi and Goodwin (2011) of research on value change showed that values can change, even substantially, and identified several mechanisms of value change. Repeated priming of values by a new environment (e.g., new organizational culture) and adaptation to life situations (e.g., new role at work) may lead to a permanent value change. As an example, a sample of Ingrian-Finns that was longitudinally followed in the migration process from Russia to Finland showed a marked increase in post-migration Universalism values (Lönqvist, Jasinskaja-Lahti, & Verkasalo, 2011). Russia being amongst the European countries scoring the lowest in Universalism values and Finland being amongst those countries scoring the highest (Verkasalo, Lönqvist, Lipsanen, & Helkama, 2009), this change may have resulted from priming and adaptation processes. Such processes of value change may also be important in the context of more purposefully directed change. Raising people's awareness of the moral dimensions of their behavior or decisions has been suggested to be an effective intervention or moral education strategy (e.g., Bazerman & Gino, 2012). This type of intervention could also be seen as causing value change: Repeatedly priming Universalism values such as Fairness and Equality should make these values more important (Bardi & Goodwin, 2011). The priming of moral values can also be pursued more generally by means of increased exposure, not only by active interventions seeking to raise moral awareness. For

instance, the organization's code of conduct, public relations management, or leadership could emphasize the importance of moral values.

Self-dissatisfaction and identification are two processes of value change that could be particularly relevant in an applied context. An interpretation of a person's rank ordering of personal values that puts the person in a negative light induces self-dissatisfaction that results in value change (e.g., Rokeach, 1973). Highlighting the more negative aspects of Conformity could thus result in a lowered adherence to Conformity values. However, at least of equal importance would be to develop a heightened adherence to Universalism values. Values develop in part through identification, becoming internalized as people take on new social identities (e.g., entering a new work group). In an organizational context, spending time with mentors (Chatman, 1991) and programs of values socialization (Chatard & Selimbegovic, 2007; Saks & Ashfort, 1997) have been shown to foster value change. Companies could also strive more strongly to attract and hire people who are intrinsically motivated to behave morally, e.g., by introducing a code of conduct to establish ethical guidelines. Unfortunately, those people scoring high in Universalism values may generally be less likely to be drawn into fields such as business (e.g., Lindeman & Verkasalo, 2005; Lönnqvist, Verkasalo, Helkama, Andreyeva, Bezmenova, et al., 2007). One reason for this lack of interest could be the high level of competition and inequality within the business community. However, it could be possible to frame certain important positions, such as mediator or compliance manager, in terms of providing opportunities to advance justice, equality, or human rights. These positions are important and should ideally be occupied by people who adhere to Universalism values, and such people could find such positions rewarding. In addition to framing the positions in terms of values that could appeal to people high in Universalism, another factor to consider in the recruitment process could be that non-monetary incentives (e.g., work environment and opportunity to make a difference) may appeal more to these people than direct monetary rewards. More generally, presenting the

company as caring about moral values, even if primarily done for marketing and strategic purposes, could, besides priming moral values and thereby fostering value change, also have the side effect of attracting morally motivated people to work for the company.

### *Limitations and Conclusions*

A general limitation of our research was that it was conducted in a laboratory setting. In an attempt to enhance the practical transferability of our results, in Study 4 we assessed a straightforward and easy to implement measure to reduce MH. Although our setting gave us more control over possible confounding factors, the question is whether and how individuals' behavior in artificial experiments is related to their behavior outside the laboratory. Some recent research on the relations between laboratory behavior and real-life behaviors or outcomes, such as the productivity of workers in manufacture (Barr & Serneels, 2009) or fishing (Carpenter & Seki, 2010), and repay of microcredit (Karlan, 2006), has suggested that laboratory behavior is associated with behavior outside of the laboratory. The relations appear to be of similar magnitude as those between responses to personality questionnaires and behavior, i.e., seldom exceeding correlations of .30. However, it may be particularly difficult to generalize laboratory results in the domain of moral behaviors – moral behaviors outside of the laboratory may be guided by stronger situational norms than other behaviors (e.g., Bardi & Schwartz, 2003). Nevertheless, there is some research showing that dictator game behavior is also generalizable to real-life contexts. For instance, the amount transferred in a dictator game has been shown to be correlated between  $r = .24$  and  $r = .40$  with actually occurring decisions on charitable giving; i.e., whether one, in conjunction with paying the compulsory university tuition fee, wants to voluntarily donate a specific amount of money to social funds (Benz & Meier, 2008). To take another example, among a sample of volunteer firefighters, dictator game giving was associated with call records that provided an objective measure of time spent volunteering as well as with

training hours (Carpenter, & Meyers, 2010). Giving in the dictator game has also been correlated with returning money that one finds in a misdirected letter (Franzen & Pointner, 2012).

More generally, money has been argued to increase the external validity of laboratory research by connecting laboratory decisions to the world outside of the laboratory (Falk & Heckman, 2009; Levitt & List, 2007). In fact, the results of Study 2 support the idea that whether behavior is monetarily incentivized to some extent determines behavior: although the majority of people claimed that they would behave morally in circumstances in which such claims were costless, less than a fifth actually behaved morally when the behavior had monetary consequences (see also Lönnqvist, Verkasalo, & Walkowitz, 2011). Whether the 10€ at stake was a large enough amount to be considered seriously by subjects is another issue. However, previous research that has explicitly tested for stake effects has found that stake size in experimental games does not generally affect the behavior of decision-makers (for a review, see Camerer, 2003). More specifically, regarding research on stake size in dictator games, investigations of games with stakes of 5\$ and 10\$ (Forsythe, Horowitz, Savin, & Sefton, 1994) and 10\$ and 100\$ (Carpenter, Verhoogen & Burks, 2005; List & Cherry, 2008) have found no differences in the distributions of offers. However, regarding the general external validity of laboratory experiments, we acknowledge that qualitative research investigating how participants experience the situation would be much needed (e.g., as an ethical dilemma or as a “research game” in which one plays to win).

Another important advantage of having participants make monetary decisions is that there is no need for deception. The gullibility of the research participants therefore does not become an issue, and there is no need to retrospectively determine whether participants actually believed in the cover story (e.g., that there was another participant involved). Most importantly, no participants have to be excluded from the data set due to suspicion regarding the research design, as is often the case in social psychological research. Including all participants eliminates

one source of potential systematic error that could distort the results.

Acknowledging the above limitations regarding the generalizability of laboratory experiments, another threat to the generalizability of our results comes from our exclusive reliance on German student samples. Although value priorities, as measured with the Portraits Values Questionnaire, tend to be extremely similar across countries (Fischer & Schwartz, 2011), nationality may still be a confounding factor: the links between values and attitudes or behaviors may vary from country to country. A recent cross-cultural meta-analysis (Boer & Fischer, 2012) found that the links between values and attitudes tended to be of average effect size in Germany, compared to other countries. It would be highly interesting to replicate our research design in more individualistic countries, such as the USA, in which the links between pro-self vs. pro-social attitudes and the corresponding values tend to be stronger or in more collectivistic countries, such as China, in which such relations tend to be weaker (Boer & Fischer, 2012; Figure 6d). Replication with non-student samples would also be a worthwhile endeavor, although prior research has tended to find highly similar results with student and representative samples (Schwartz & Rubel-Lifschitz, 2009).

Another potential limitation was that we always started the sessions with dictator games and concluded them by gathering ratings of personal values. We always kept this order to keep the dictator game decisions unbiased; completing a values questionnaire could activate personal values and thereby influence subsequent behavior (e.g., Verplanken & Holland, 2002). However, it is also possible that one's behavior in the dictator games could influence one's ratings of values, thereby inflating the size of the observed correlations. However, in a different data set, we noted that Universalism values collected in an online web survey and Universalism values collected in the laboratory right after playing a social dilemma game were equally strongly associated with ethical behavior in the social dilemma game (Spearman  $\rho_s(23) = .28$  and  $.26$ ,

respectively). Furthermore, in yet another data set, the correlations between Universalism values and social dilemma decisions were actually stronger if the values measure was administered before instead of after the task (Pearson  $r(43) = .31$  and Pearson  $r(52) = .17$ , respectively). These data suggest that the predictive power of personal values may generally not have been heightened by being measured after the dictator game decisions.

Finally, it should be acknowledged that although our results indicate that misrepresentations of simple and discrete behaviors may be motivated merely by wanting to appear good, more self-deceptive processes may be at work regarding more abstract or complex behaviors that may be easier to misrepresent. For instance, although research has shown only very few and very weak relations between environmental attitudes and ecological behavior (e.g., Bamberg, 2003), people reporting environmentally friendly attitudes may honestly, although self-deceptively, believe that they tend to behave in an environmentally friendly way. Even in the present research design, we cannot exclude the possibility that self-deception may have occurred after participants had decided to flip the coin. However, our results do suggest that self-deception was not the driving force behind such a decision.

Perhaps the central contribution of research on moral hypocrisy is that it reveals that moral failures are often due to a lack of moral motivation; i.e., people are primarily motivated to appear moral and not to be moral. Batson and colleagues have investigated and rejected alternative accounts of moral hypocrisy, such as misperception of hypocritical behavior as moral (Batson et al., 1999; Study 1) and low salience of social standards (Batson et al., 2002; Study 1). We add to this literature by showing that moral hypocrisy is not self-deceiving. Rather, moral hypocrisy appears to be motivated only by the desire to appear fair in the eyes of others.

Another contribution of the present research is that we position both moral integrity and hypocrisy into the framework provided by Schwartz' (1992) values theory. Although there is a plethora of other individual difference variables that we could have focused on in this context,



such as manipulateness, self-monitoring, or communal narcissism, personal values could be particularly useful in research on morality because they appear to cover both social preferences and social signaling as motives of moral behavior. Whereas moral integrity is motivated by Universalism values, moral hypocrisy is driven by adherence to Conformity values. Connecting moral integrity and moral hypocrisy to established and well-researched constructs such as personal values should illuminate promising avenues for future research.

Footnotes.

1. Our characterization of directly choosing the 8/2 allocation as selfish could be disputed – the 8/2 allocation could be considered perfectly justified if one interprets the situation as one in which the person assigned the role of the dictator has already won (by being randomly assigned the dictator role) and should now be free to enjoy the rewards of the victory without feeling guilty. To investigate how the three different choices were actually evaluated by our participants, we asked a highly similar but completely independent sample of students ( $N = 126$ ) to rate choosing (a) 5/5, (b) 8/2, and (c) the coin flip, on three bipolar 9-point scales. The scales were “unfair-fair”, “immoral-moral”, and “bad-virtuous”, and the midpoint of each scale (response option 5) was marked “neutral”. The Cronbach’s alphas for these three-item scales were, for evaluations of 5/5, 8/2, and the coin flip, .56, .75, and .85, respectively. The means of these scales were 8.09 (SD = 1.01), 3.56 (SD = 1.66), and 6.27 (SD = 2.09), respectively. All of the means differed from each other (in paired-samples  $t$ -tests, all  $t > 9.60$ , all  $p < .001$ ), and all of the means differed from the midpoint of the scale (in one-sample  $t$ -tests against the value of five, all  $t > 6.87$ , all  $p < .001$ ). These ratings thus suggest that choosing the 8/2 allocation was perceived by students as unfair or immoral, and we accordingly refer to this choice as selfish. By contrast, letting the coin decide, or better yet, directly choosing the 5/5 allocation, was perceived as fair.

## References

- Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, 50, 179-211.
- Bartling, B., & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies*, 79, 67-87.
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, D. W. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335–1348.
- Batson, C. D., Lishner, D. A., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E. L., et al. (2003). “. . . As you would have them do unto you”: Does imagining yourself in the other’s place stimulate moral action? *Personality and Social Psychology Bulletin*, 29, 1190-1201.
- Batson, C. D., Thompson, E. R., & Chen, H. (2002). Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology*, 83, 330–339.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, 77, 525–537.
- Bamberg, S. (2003) How does environmental concern influence specific environmentally related behaviours? A new answer to an old question. *Journal of Environmental Psychology*, 23, 21–32.
- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society of London, Series B*, 274, 749–753.

- Bardi, A. & Goodwin, R. (2011). The dual route to value change: individual processes and cultural moderators. *Journal of Cross-Cultural Psychology*, 42, 271-287.
- Bardi, A., & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin*, 29, 1207-1220.
- Barr, A., & Serneels, P. (2009). Reciprocity in the workplace. *Experimental Economics*, 12, 99-112.
- Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, 8, 85-94.
- Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field? Evidence from donations. *Experimental Economics*, 11, 268–281.
- Boer, D. & Fischer, R. (in press). How and when do personal values guide our attitudes and sociality? Explaining cross-cultural variability in attitude-value linkages. *Psychological Bulletin*.
- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Carpenter, J., & Meyers, C. K. (2010). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics*, 94, 911–920.
- Carpenter, J., & Seki, E. (2010). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay". *Economic Inquiry*, 49, 612–630.
- Carpenter, J., Verhoogen, E., Burks, S., 2005. The effect of stakes in distribution experiments. *Economics Letters*, 86, 393–398.
- Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in

- dictator and ultimatum games. *Journal of Economic Behavior and Organization*, 68, 29-35.
- Chatard, A., & Selimbegovic, L. (2007). The impact of higher education on egalitarian attitudes and values: Contextual and cultural determinants. *Social and Personality Psychology Compass*, 1, 541-556.
- Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36, 459-484.
- Chemers, M.M. (1993). An integrative theory of leadership. In M.M. Chemers & R. Ayman (Eds.), *Leadership theory and research: Perspectives and directions* (pp. 293-320). San Diego, CA: Academic Press.
- Chemers, M.M. (1997). *An integrative theory of leadership*. Mahwah, NJ: Lawrence Erlbaum.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Denissen, J. J. A., Geenen, R., Selfhout, M., & van Aken, M. A. G. (2008). Single-item Big Five ratings in a social network design. *European Journal of Personality*, 22, 37-54.
- Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170-177.
- Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326, 535 - 538.

- Fernández-Dols, J.M., Aguilar, P., Campo, S., Vallacher, R., Janowsky, A., Rabbia, et al. (2010). Hypocrites or maligned cooperative participants? Experimenter induced normative conflict in zero-sum situations. *Journal of Experimental Social Psychology*, *46*, 525-530.
- Fehr, E., Fischbacher, U. & Gächter, S. (2002) Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, *13*, 1–25.
- Fehr, E., & Schmidt, K.M. (1999). A theory of fairness, competition and co-operation. *Quarterly Journal of Economics*, *114*, 817–868.
- Fenigstein, A., & Vanable, P. A. (1992). Paranoia and self-consciousness. *Journal of Personality and Social Psychology*, *62*, 129–138.
- Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of deindividuation in a group. *Journal of Abnormal and Social Psychology*, *47*, 382–389.
- Fishbach, A., Zhang, Y., & Koo, M. (2009). The dynamics of self-regulation. *European Review of Social Psychology*, *20*, 15–344.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*, 171-178.
- Fischer, R. & Schwartz, S. H. (2011). Whence differences in value priorities? Individual, cultural, or artifactual sources. *Journal of Cross-Cultural Psychology*, *42*, 1127-1144.
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347–369.
- Franzen, A. & Pointner, S. (in press). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S. P., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*, 366–385.

- Griffin, D., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 319–359). San Diego, CA: Academic Press.
- Haidt, J. (2003). The moral emotions. In: Davidson, R. J., Scherer, K. R., Goldsmith, H. H., (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.
- Helkama, K. (2004) Values, role-taking and empathy in moral development. *New Review of Social Psychology*, 3, 103-111.
- Helkama, K. (2011). Social psychology of morality and moral development: A functional system approach. In J. P. Valentim (Ed.), *Societal approaches in social psychology* (pp. 131-149). Frankfurt am Main: Peter Lang GmbH Company.
- Karlan, D. (2006). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95, 1688-99.
- Ketelaar, T., & Au, W. T. (2003). The effects of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction . *Cognition and Emotion* , 17 : 429 – 453
- Kohlberg, L. (1984). *The psychology of moral development*. San Fransisco, Harper & Row.
- Keller, J. and Pfattheicher, S. (2011). Vigilant self-regulation, cues of being watched and cooperativeness. *European Journal of Personality*, 25, 363–372.
- Leikas, S., Lönnqvist, J.-E., Verkasalo, M., & Lindeman, M. (2009). Regulatory focus systems and personal values. *European Journal of Social Psychology*, 39, 415-429.
- Levitt, S. D. & List. J. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21, 153-174.
- Lindeman, M., & Verkasalo, M. (2005). Measuring values with the Short Schwartz's Value

Survey. *Journal of Personality Assessment*, 85, 170-178.

- List, J. A., & Cherry, T. L. (2008). Examining the role of fairness in high stakes allocation decisions. *Journal of Economic Behavior & Organization*, 65, 1–8.
- Lönnqvist, J.-E., Jasinskaja-Lahti, I., & Verkasalo, M. (2011). Personal values before and after migration: A longitudinal case study on value change in Ingrian–Finnish migrants. *Social Psychological and Personality Science*, 2, 584-591.
- Lönnqvist, J.-E., Leikas, S., Paunonen, S., Nissinen, V., & Verkasalo, M. (2006). Conformism moderates the relations between values, anticipated regret, and behavior. *Personality and Social Psychology Bulletin*, 32, 1469–1481.
- Lönnqvist, J.-E., Verkasalo, M., Helkama, K., Andreyeva, G. M., Bezmenova, I., Manganelli Rattazzi, A. M., Niit, T., & Stetsenko, A. (2007). Self-esteem and values. *European Journal of Social Psychology*, 39, 40-51.
- Lönnqvist, J.-E., Verkasalo, M., Wichardt, P., & Walkowitz, G. (2012). *Personal Values and Ethical Behavior in Strategic Interactions: Distinguishing Value-Expressive from Value-Ambivalent Behaviors*. Unpublished manuscript.
- Lönnqvist, J.-E., Verkasalo, M., & Walkowitz, G. (2011). It pays to pay – Big Five personality influences on co-operative behavior in an incentivized and hypothetical prisoner’s dilemma game. *Personality and Individual Differences*, 50, 300-304.
- Lönnqvist, J.-E., Walkowitz, G., Verkasalo, M., & Wichardt, P. (2011). Situational power moderates the influence of Self-Transcendence vs. Self-Enhancement values on behavior in ultimatum bargaining. *Journal of Research in Personality*, 45, 336-339.



- Lönnqvist, J.-E., Walkowitz, G., Wichardt, P., Lindeman, M., & Verkasalo, M. (2009). The moderating effect of conformism values on the relations between other personal values, social norms, moral obligation, and single altruistic behaviours. *British Journal of Social Psychology, 48*, 525–546.
- Maio, G. R. (2010). Mental representations of social values. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 42). New York, NY: Academic Press.
- Maio, G. R., & Olson, J. M. (1995). Relations between values, attitudes, and behavioural intentions: The moderating role of attitude function. *Journal of Experimental Social Psychology, 31*, 266-285.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*, 28–50.
- Mele A. (1987). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press.
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4/5*, 344-357.
- Monin, B., & Merritt, A. (2011). Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In M. Mikulincer & P. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil. Herzliya Series on Personality and Social Psychology* (Vol. 3). American Psychological Association.
- Myyry, L., & Helkama, K. (2002). The role of value priorities and professional ethics training in moral sensitivity. *Journal of Moral Education, 31*, 35–50.

- Myyry L, Juujärvi S, Pessa K (2010). Empathy, perspective taking and personal values as predictors of moral schemas. *Journal of Moral Education*, 39, 213-233.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598 – 609.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49 –69). Mahwah, NJ: Erlbaum.
- Rest, J. (1979). *Development in Judging Moral Issues*. University of Minnesota Press.
- Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Saks, A. M., & Ashforth, B. E. (1997). Organizational socialization: Making sense of the past and present as a prologue for the future. *Journal of Vocational Behavior*, 51, 234–279.
- Scheepers, D., Spears, R., Manstead, A. S. R., & Doosje, B. (2009). The influence of discrimination and fairness on collective self-esteem. *Personality and Social Psychology Bulletin*, 35, 505-515.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In Zanna, M. (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, pp, 1-65). New York: Academic Press.
- Schwartz, S. (1977). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, 10, 221-279. San Diego, CA: Academic Press.
- Schwartz, S. H. (2009). Culture matters: National value cultures, sources and consequences. In C.-Y. Chiu, Y.Y. Hong, S. Shavitt, & R. S. Wyer, Jr. (Eds.), *Understanding culture: Theory, research and application*. New York: Psychology Press.

- Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology, 32*, 519–542.
- Schwartz, S., H. & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of Personality and Social Psychology, 97*, 171-185.
- Schwartz, S. H., & Tessler, R. C. (1972). A test of a model for reducing measured attitude-behavior discrepancies. *Journal of Personality and Social Psychology, 24*, 225-236.
- Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology, 36*, 3–18.
- Schweitzer, M. E., Hershey, J. & Bradlow, E. (2006) Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*, 1–19.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “big three” of morality (autonomy, community, and divinity), and the “big three” explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119 –169). New York: Routledge.
- Stouten, J., De Cremer, D. & van Dijk, E. (2006) Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin, 32*, 894 – 906.
- Shalvi, S., Dana, J., Handgraaf, M. J. J. & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*, 181-190.
- Shu, L., Mazar, N., Gino, F., Ariely, D., Bazerman M. 2012. Signing at the beginning makes

ethics salient and decreases dishonest self-reports in comparison to signing at the end,

*Proceeding of the National Academy of Science, 109*, 15197-15200.

Silfver, M., Helkama, K., Lönnqvist, J.-E., Verkasalo, M. (2008) The relation between value priorities and proness to guilt, shame and empathy. *Motivation and Emotion, 32*, 69–80.

Sugiyama, L. S., Tooby, J., Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences, 99*, 11537–45.

Susewind, M. Hoelzl, E. & Walkowitz, G. (2012). *To be Moral you need to be seen: Lack of Social Recognition encourages further Moral Striving*. Submitted manuscript.

Tetlock, P. E., & Manstead, S. R. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review, 92*, 59–77.

Treviño, L. T., Weaver, G., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management, 32*, 951–990.

Trivers, R. (1985) Deceit and self-deception. In: *Social evolution*, pp. 395–420. Benjamin/Cummings.

Trivers, R. (1991) Deceit and self-deception: The relationship between communication and consciousness. In: *Man and beast revisited*, ed. M. Robinson & T. L. Tiger, pp. 175–91. Smithsonian Press.

von Hippel W., & Trivers R. (2011). The evolution and psychology of self-deception [Target article and commentary]. *Behavioral & Brain Sciences, 34*, 1–56.

Verkasalo, M., Lönnqvist, J.-E., Lipsanen, J., & Kelkama, K. (2009). European norms and equations for a two dimensional presentation of values as measured with

- Schwartz's 21-item Portrait Values Questionnaire. *European Journal of Social Psychology*, 39, 780–792.
- Verplanken, B., & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology*, 82, 434–447.
- Vorauer, J. D., & Ross, M. (1999). Self-awareness and feeling transparent: Failing to suppress one's self. *Journal of Experimental Social Psychology*, 35, 415–440.
- Waples, E. P., Antes, A. L., Murphy, S. T., Connelly, S., & Mumford, M. D.: 2008. A meta-analytic investigation of business ethics instruction'. *Journal of Business Ethics*, 87, 133–151.
- Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, 21, 311–314.
- Zimbardo, P. (1969). The human choice: Individuation, reason, and order vs. deindividuation, impulse, and chaos. In W.J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 17, pp. 237–307). Lincoln: University of Nebraska Press.

Table 1. Dictator Behavior in Studies 1, 2, and 3.

	Outcome 8/2		Outcome 5/5	
Study 1	Without coin	With coin	Without coin	With coin
Watched	12 (38%)	16 (50%)	4 (13%)	0
Unwatched	18 (56%)	10 (31%)	4 (13%)	0
Collapsed	30 (47%)	26 (40%)	8 (13%)	0
Study 2				
Incentivized	26 (81%)		6 (19%)	
Hypothetical	15 (47%)		17 (53%)	
Study 3	22 (69%)	2 (6%)	6 (19%)	2 (6%)
Study 4	14 (44%)	7 (22%)	7 (22%)	4 (12%)

Table 2. Regression Analyses predicting the Distributively Fair Decision in Studies 1, 2, and 3.

	Study 1		Study 2		Study 3		Study 4	
	Universalism (U)	Conformity (C)	Universalism (U)	Conformity (C)	Universalism (U)	Conformity (C)	Universalism (U)	Conformity (C)
Age	.01	.06	-.05	-.04	.20	.13	-.29	-.15
Sex	-.27	-.14	.26*	.27*	.08	.10	.00	.08
Manipulation	.42	.83	-.68	-.61	NA	NA	NA	NA
Value (U or C)	.45*	.22	-.09	-.06	.41*	.18	.34	-.14
Value x Manipulation	-.50	-.95	1.08	.99*	NA	NA	NA	NA

*Note.* All table entries are Beta values from regression analyses with all predictor variables entered. In Studies 1 and 2, Manipulation refers to the Watched vs. Unwatched and the Hypothetical vs. Incentivized conditions, respectively. Study 3 did not include a manipulation.

FIGURE CAPTIONS

Figure 1. Equal vs. selfish choice regressed on low and high Conformity scores in the hypothetical and incentivized condition (Study 2).

*Note:* Low score = 1 SD below the mean; high score = 1 SD above the mean.



**APPENDIX. Instructions for the Dictator Game (Study 1)****Description of the 1st part of the experiment**

This experiment is about the **distribution of an amount of 10€** between **two people**, who are referred to as **Person A** and **Person B**. Person A can determine how the amount is to be distributed. Person A is provided with a **coin** at his/her place, which can be used to determine the distribution. The coin has one side with the distribution option **"8/2"** and another side with the distribution option **"5/5"**.

**Person A has three options to distribute the 10€.**

- **Person A can choose the distribution "8/2" directly. Person A gets 8€ and Person B 2€.**
- **Person A can choose the distribution "5/5" directly. Person A gets 5€ and Person B 5€.**
- **Person A can use the coin to determine which distribution is to be realized.**

**Person B does not make a distribution decision / decision concerning the distribution.** He/she gets the amount that results from the distribution decision of Person A. **However, Person B does not learn/learns whether Person A used the coin for his/her decision.**

At the beginning of the experiment, **your role** is determined **randomly**. Furthermore, you are **matched randomly with another participant** who decides in the other role.

**How is the payment calculated?**

- If **Person A** chooses the **distribution "8/2"** directly: **Person A gets 8€. Person B gets 2€.**
- If **Person A** chooses the **distribution "5/5"** directly: **Person A gets 5€. Person B gets 5€.**
- If Person A chooses **the coin to determine the distribution**: The payments of Person A and B **depend on the coin flipping**. **Person A** gets the amount that **comes first** on the coin. **Person B** gets the amount that **comes second**.

Figure

