Journal of Experimental Social Psychology 46 (2010) 525–530

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Hypocrites or maligned cooperative participants? Experimenter induced normative conflict in zero-sum situations

Jose-Miguel Fernandez-Dols [a,*], Pilar Aguilar [a], Silvia Campo [a], Robin R. Vallacher [b], Alisha Janowsky [c], Hugo Rabbia [d], Silvina Brussino [d], Melvin J. Lerner [e]

[a] Facultad de Psicología, Universidad Autónoma de Madrid, c/ Ivan Pavlov 6, 28049 Madrid, Spain
[b] Department of Psychology, Florida Atlantic University, Boca Raton, FL 33431-0991, USA
[c] Department of Psychology, University of Central Florida, 4000 Central Blvd. Orlando, FL 32816-1390, USA
[d] Facultad de Psicología, Universidad Nacional de Córdoba, Ciudad Universitaria 5000, Córdoba, Argentina
[e] 4460 NW 27th Avenue, Boca Raton, FL 33434, USA

ARTICLE INFO

Article history:
Received 14 May 2009
Revised 30 January 2010
Available online 6 February 2010

Keywords:
Moral hypocrisy
Self-interest
Justified self-interest
Supererogation
Social justice
Social norms
Morality
Normative conflict

ABSTRACT

The failure to recognize the influence of two distinct forms of moral norms can lead to the misattribution of moral behavior to egoistic motives. This is illustrated in the research of Batson and his colleagues (e.g., Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997). They reported the appearance of moral failure and hypocrisy motivation in several experiments employing essentially the same "zero-sum" experimental situation. They cited as evidence the discrepancy between participants' apparently self-serving private acts and their subsequent public ratings of the morality of what they had done as well as their recognition of the "most" moral way to behave. The research reported here supported an alternative explanation that located the experimenter's implicit and explicit instructions as the source of the discrepancy between the participants' private acts and their public ratings. The findings confirmed the hypothesis that Batson and his colleagues had not merely made moral norms "salient". They had actually presented their participants with contradictory "demands": explicitly inviting them to meet the norm of justified self-interest in private but then give public lip-service to the experimenter's instructions as to a supererogatory way to behave. When either of the demands was removed, the "hypocrisy" no longer occurred.

It is commonly recognized that people maintain a moral view of themselves and those around them. Anyone who is culturally conversant recognizes that there are explicit standards of conduct that are to be maintained and socially valued. There are moral norms, closely associated with the various statuses in the social structure, that typically define obligations and privileges, i.e., how people are supposed to behave in various contexts. The failure to meet those expectations is considered to be a moral failure, and merits varying degrees of social sanction. On the other hand fulfilling one's obligations does not merit moral accolades. Quite the opposite is true of those moral norms that define virtuous acts and self sacrifice in order to benefit others i.e., "above and beyond the call of duty". These kinds of supererogatory acts naturally accrue special praise and moral credits. At the same time, it is also recognized that the failure or refusal to engage in these especially meritorious acts is not considered a disgrace or a moral failure. Most people are very willing to praise the Mother Teresa's and soldiers who throw themselves on a grenade to save the lives of their fellow soldiers, but they fully realize that there is no shame in not volunteering for such self sacrificial acts.

The distinction between these two kinds of moral norms has a long history in Western thought (Heyd, 2006). For example, since the Middle Ages, a long debate focused on the theological differences between no-sin and perfection (e.g., taking a wife versus opting for virginity) or, in secular philosophical terms, between actions that are in compliance with the moral requirements of a particular situation (meeting one's obligations) and actions that are most moral but not required (the "good") (e.g., paying your taxes versus donating all your fortune to a good cause). Supererogation is good because it advances behavior out of love and altruistic impulses; it is saintly or heroic and it merits moral accolades but it is not governed by norms of justice and rights. Legitimate acts are governed by norms of justice and rights, i.e., obligations and privileges, but their accomplishment does not merit moral accolades.

The failure to recognize the distinct ways that these two kinds of norms enter into people's lives can lead to misleading interpretations of how people manifest moral considerations in their social acts. An important example of this has appeared in the literature intended to examine the appearance of an important moral failure: hypocrisy.

* Corresponding author. Address: Facultad de Psicología, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
E-mail addresses: fernanjj@bc.edu, jose.dols@uam.es (J.-M. Fernandez-Dols).

0022-1031/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2010.02.001

Batson, Thompson, and Chen (2002), Batson, Thompson, Seuferling, Whitney, and Strongman (1999) and Batson et al. (1997) have provided the most representative and influential approach to moral hypocrisy in contemporary social psychology in a series of laboratory experiments. Their aim was to present their participants with a clear conflict between either choosing to benefit them or adhering to a more costly moral principle. According to Batson many of their participants revealed failures in moral integrity by choosing to benefit themselves while acting in private even though they reported awareness that benefiting themselves – rather than being generous to others – was not the "most moral" way to behave. And then, many of these participants completed the act of hypocrisy by publicly rating their patently selfish behaviors as relatively moral (see also Valdesolo & DeSteno, 2007, 2008).

There are, however, considerable reasons to question these conclusions. The first question that will be addressed by the research presented here is whether or not Batson et al. actually created a conflict between the participants pursuing their own egoistic interests or complying with their moral standard? And second, did the participants who elected the more desirable outcome for themselves in private reveal a failure in moral integrity? And finally, did their subsequent public ratings of their private acts as relatively moral reveal hypocritical motivation?

The answers to these important questions can be approached by first examining the conceptual basis of Batson's and colleagues approach. Upon careful analysis it seems clear that the theoretical rationale underlying Batson's research and subsequent inferences did not take into account the important distinction between legitimate versus supererogatory moral norms and actions.

In Batson's experimental paradigm the participants were led to believe that they and another student were paired for purposes of the experiment. They learned there were two tasks that differed considerably in desirability: one was portrayed as a pleasant task with the opportunity to earn raffle tickets for a prize, whereas the other was portrayed as a rather dull and boring task with no possibility of earning raffle tickets: a classic zero-sum situation. Then, the participants were told they had been selected to decide which of them would be assigned to the desirable and undesirable conditions.

It is important to recognize, at this point, that earlier research involving this kind of zero-sum situation had produced evidence for the deserving-norm that was then termed the "norm of justified self-interest" (Lerner, 1971; Lerner & Lichtman, 1968; Lerner, Miller, & Holmes, 1976). That research demonstrated that in the absence of prior claims to the desired outcome, the norm of "justified self-interest" applies in situations of direct or parallel competition over desired, and, especially, indivisible resources. The participants in that research recognized that under those conditions, it is entirely appropriate for people to take advantage of every opportunity to gain the desired resource, as long as they do not violate explicit rules of fair competition (Lerner, 1971; Lerner & Lichtman, 1968; Lerner et al., 1976). As a consequence, the participants felt they could legitimately act to benefit themselves as long as they did nothing that was explicitly prohibited.

To validly interpret participants' reactions in this zero-sum context it is necessary to know whether, as Batson et al. claim, the participants' were actually confronted with a relatively simple conflict between taking the desired resource for themselves and adhering to their moral standard; or, did they find themselves facing a conflict between meeting the experimenter's two conflicting expectations, "demands", as to how they, as cooperative participants, were supposed to react?

Batson's paradigm

Batson et al. (1997, 1999, 2002) experimenters explicitly assured their participants that: "The decision is entirely up to you. You can assign yourself and the other participant however you choose". Immediately after those initial instructions, the experimenters then introduced clearly contradicting instructions that were intended to make the moral standard of "fairness" "salient":

> Most participants feel that giving both people an equal chance – as by, for example, flipping a coin – is the fairest way to assign themselves and the other participant to the tasks (we have provided a coin for you to flip if you wish) (Batson et al., 1997, p. 1341).

Or, to make altruism "salient", Batson et al. (1999) added:

> Most participants feel that giving the other person the positive-consequences task is the most morally right way to assign themselves and the other participant to the tasks (p. 532).

Having been assured privacy – and that the experimenter would inform the other participant that the assignments was purely by chance – participants were then left alone to make their choices. After that, they were given various questions to answer and rate alternatives concerning their views of the morality of various responses and the morality of their own behavior. Those were the obviously public reactions to be used for purposes of comparison.

How could the participants' be expected to respond to these two conflicting messages? They could have ignored one of them: either the experimenters' having legitimized their taking the desirable condition for themselves, or their thinly disguised subsequent moral injunctions. Instead, a significant number of them gained the experimenter's label of hypocrite by responding to both: they chose the desirable condition for themselves in private, and then gave public lip-service to the experimenters' recommended moral standard.

That raises the question of whether or not it is valid to portray the participants' reactions as moral failures followed by acts of hypocrisy. The answer depends in great part on whether or not the experimenters' instructions merely made the moral norm "salient" as claimed by Batson et al. (1999), or, as suggested here, those instructions effectively communicated the experimenters' own recommendation and preferences with which cooperative participants did their best to comply, or, at least, not openly contradict.

The research described here was designed to gather evidence concerning whether or not Batson's instructions both legitimized the participants' acting according to a legitimate norm of justified self-interest, and then – rather than merely making the supererogatory, most moral norm "salient" – the experimenters' instructions actually recommended the participants use a coin flip to determine the assignments.

Our experiment

The main hypothesis is that – rather than first pursuing, then disguising their self-interest, cooperative participants would comply with whatever norms are defined as appropriate by the experimenters, both in their decisions and in their public ratings. In this situation, the experimenters' instructions confirmed following the norm of justified self-interest and/or the moral norm of altruistic generosity. However, when both norms were introduced, it became more socially desirable to publicly endorse the moral norm of generosity or fairness while following the norm of justified self-interest in private, especially if the experimenter had previously modeled that pattern.

The strategy employed to examine these hypotheses involved replicating the Batson et al. (1997, 1999) procedure to elicit the participants' "moral failures" and "hypocrisy." Then, by systematically altering either of the two critical elements in the experimenters'

instructions, it should be possible to reveal the participants' adherence to either the norm of justified self-interest or their compliance with the random assignment-fairness instructions. In addition, the presence of experimenter demands in the Batson et al.'s "salience" instructions could be documented by simply making generosity, fairness, and justified self-interest norms salient without endorsing any one of them as "most fair." If the "salience" instructions employed by Batson et al. did not serve as an experimenter demand, then the participants' responses in this condition should closely resemble those in the Batson-replication condition.

To test these hypotheses, we randomly assigned participants to one of four experimental conditions:

Condition 1 (Conflicting Norms): This condition replicated the procedure Batson employed to make salient the rule of the participants allowing chance to decide the assignment of tasks[1] by flipping a coin (Batson et al., 1997).

Condition 2 (Moral Norm): In this condition, we deleted the instructions to the recipient that the decision of who will be assigned to which condition "is entirely up to you. You can assign yourself and the other participant however you choose ... your anonymity is assured". What remained were the instructions that Batson described as making the moral norm salient: "Most participants feel that giving both people an equal chance is the fairest way to assign themselves and the other participant to the task (you have a coin if you wish)" (Batson et al., 1997). By removing the instructions that explicitly entitled participants to pursue their own self-interest, we expected that they would follow the experimenter's implicit instructions to flip the coin to make the assignments according to the outcome, even though they were acting in private.

Condition 3 (Discretionary Norm): A third condition was employed to make three norms equally salient without the experimenter appearing to endorse any one in particular. After giving participants the opportunity to assign the conditions, the experimenter accurately informed them that some previous participants had assigned themselves to the desirable condition, some had assigned the other to the desirable condition, and some had left the assignments up to chance by flipping a coin. If it is true that simply making a moral norm salient is sufficient to influence participants' behavior and beliefs about the most moral way to behave, then the participants in this condition should express allegiance to these norms in their behavior and descriptions of the most moral way to act. If, however, Batson's (e.g., Batson et al., 1999) "salience" induction actually instructed the participants concerning the experimenter's moral preferences, then in this condition the accurate descriptions of previous participants behaviors that included flipping a coin, as well as simply assigning the desirable outcome either to oneself or the other participant, should leave the participants free to follow the initial explicitly introduced norm of justified self-interest. As a result, they would feel entitled to elect the desirable condition for themselves.

Condition 4 (Obedient Cooperation): A fourth condition provided participants with the opportunity to indulge their preference for the more desirable condition in private after they had been told that random assignment was required for the scientific purposes of the experiment. In the absence of the instructions entitling the participants to act in their own best interests, we expected that the participants would be cooperative participants and adopt the random assignment in private. Also, since they were merely meeting their obligations, absent the experimenters moral recommendations, they would be less likely to see their flipping the coin as relatively meritorious whether that eventuated in their assigning the desirable condition to themselves or the other.

As for the participants' public statements and their self-ratings of the morality of their decisions, we expected to observe little evidence of "hypocrisy", with the exception of the Conflicting Norms condition in which the norm of entitlement conflicted with the experimenter's moral injunction. Of course, participants should recognize, when asked, that it is more virtuous and noble not to take advantage of one's good fortune, and instead express allegiance to the idea of being fair to all concerned, especially after being told that by the experimenter. Such recognition, however, does not undermine the moral acceptability of following the norm of justified self-interest in their covert behavior. We also expected the public ratings of the morality of their own choices to reflect compliance with the experimenters' moral recommendations.

As in Batson's original studies, in order to assess their hypocrisy the participants' were then given a questionnaire to complete that included their reactions to making their decision, as well as their ratings of the morality of their decisions ("Do you think the way you made the task assignment was morally right?").

Method

A sample of 127 Spanish Biology students, and Argentinean Social Sciences students who had not participated in psychology experiments volunteered for this experiment. After the post-experimental debriefing, 23 participants were excluded because they expressed doubts about the veracity of the experimental situation or specific guesses about the goals of the study. Eventually, 104 participants (36 males, 68 females, mean age = 21.87) were included in the final sample (61 Spaniards and 43 Argentineans). Over all, the assignation of participants to cells was slightly uneven (30, 29, 22 and 23 participants per condition) because of coordination problems between the two laboratories. Nevertheless, the distribution per condition in the Spanish sample was practically uniform (15 participants per cell, excluding a cell with 16 participants) and the results were identical to those obtained in the total sample. One experimenter from each lab who was unaware of the specific condition ran at the time, applied the protocol in the two countries; the protocol was a replication of Batson et al.'s (1997, 1999) protocols and debriefings.

In order to detect how the participants made their assignments, as in Batson, Thompson, and Chen's (2002) Study 2, an invisible micro-camera recorded the participants' behavior during the experiment.[2] That allowed a direct comparison of the participants' private acts with their subsequent public ratings.

Results

Participants' assignments

Following Batson et al. (2002, Study 2), we classified the task assignment choices into four categories: "Assign other to positive task", "flip the coin, win and assign self to positive task", "not flip the coin, assign self to positive task", and "fiddle the coin flip, assign self to positive task".

Whereas criteria for inclusion for the three first categories are self-evident, Batson et al.'s criteria of inclusion for the last category were more complex. The category includes those who flipped the coin and got the positive task for the other and assigned the positive task for themselves but also those who did not flip the coin at all, kept the positive task for themselves and reported, in the

[1] One choice was portrayed as a pleasant task with the opportunity to earn raffle tickets for a prize, whereas the other was portrayed as a rather dull and boring task with no possibility of earning raffle tickets. The winner of the raffle was rewarded with the equivalent of 20 Euros in Argentinean pesos (at the currency exchange rate of the date) or 50 Euros. Currency amounts were considered roughly equivalent in attractive across countries.

[2] The coin sides were painted in different colors; the camera clearly recorded the outcome of flipping the coin.

Table 1
Experiment 1: mean rated morality of the way participants assigned the tasks following Batson, Thompson and Chan's task assignment behavior categories.

Task Assignment Behavior Category	Condition															
	Conflicting Norms				Moral Norm				Discretionary Norm				Obedient Cooperation			
	N	%	M	SD	N	%	M	SD	N	%	M	SD	N	%	M	SD
1. Assign other to positive task	5	18	8.60_{a}	.54	12	43	7.92_{ab}	1.44	6	27	7.33_{a}	1.30	9	45	7.11_{a}	2.71
2. Flip the coin, win and assign self to positive task	6	21	8.33_{a}	1.21	7	25	8.71_{a}	.48	2	9	7.00_{a}	2.82	11	55	6.27_{a}	3.69
3. Not flip the coin, assign self to positive task	8	29	2.75_{b}	1.98	6	21	4.17_{c}	2.63	11	50	5.45_{a}	3.11				
4. Fiddle the coin flip, assign self to positive task	9	32	6.00_{a}	3.12	3	11	5.33_{bc}	4.04	3	14	8.33_{a}	1.15				

Note: means in the same column that do not share subscripts differ at $p < .05$.

post-experimental questionnaire or the post-experimental interview, that they flipped the coin and won the positive task for themselves. For ethical and methodological reasons (we did not want the participants to be publicly exposed by openly asking them about their behavior and informing them, short afterwards, that we had a video record of their actual behavior) we adapted Batson et al.'s criterion of inclusion into the "fiddle the coin flip, assign self to positive task": we included those who did not flip the coin and kept the positive task for themselves but spontaneously claimed to have flipped the coin after the experiment or clearly implied to do so in the post-experimental questionnaire (by claiming that the most moral way of assigning the task was random, and rating both "the way you made your own assignation" and "flipping the coin" as at least twice as moral than keeping the task for yourself). These results are comparable to those reported by Batson et al. (2002, Study 2).[3]

As expected (see Table 1), the proportion of participants who assigned the positive task to themselves but did not get this outcome by flipping the coin reflected the norms that the experimenters recommended and legitimized. That proportion was significantly higher in Batson et al.'s (1997) original condition replication (Conflicting Norms condition: 17 vs. 5) than in two conditions that did not explicitly legitimize the norm of self-interest: Moral Norm: (9 vs. 12; Chi Square (1, $N = 43$) = 5.32, $p = .02$, two tails) and Obedient Cooperation (0 vs. 9; Chi Square (1, $N = 29$) = 13.07, $p < .01$, two tails). When self-interest was legitimized as one of the possible options along with altruism and random assignment (Discretionary Norm), the proportion of those who kept the positive task for themselves (and did not flip the coin) was similar to Batson et al.'s original condition replication (14 vs. 6; Chi Square (1, $N = 41$) = .71, $p = 79$).

Participants' ratings

Participants' public morality ratings also reflected their awareness of the experimenters' preferences.

Reflecting the experimenters "salience" instructions (i.e., moral recommendations), the participants' public morality ratings of random assignment in the Conflicting Norms ($M = 7.90$) and the Moral Norm conditions ($M = 7.03$) were significantly higher than were the ratings in those conditions without the "salience" instructions: the Discretionary Norm ($M = 5.50$) and Obedient Cooperation ($M = 6.65$) conditions (7.47 vs. 6.09, $t(78.45) = 2.49$, $p = .01$). This finding supports the argument that the experimenter's instructions in the first two conditions that were purported to make the participants own moral norm of random assignment "salient", actually communicated a recommendation and preference – rather than simply making the participants aware of the norm of random assignment by flipping the coin.

A 2 (conditions 1 and 2 with a "salient-recommended" Moral Norm vs. conditions 3 and 4, no "salient-recommended" moral norm) × 4 (Task Assignment Behavior Category, see Table 1) ANOVA yielded a significant main effect of the participants' task assignment decision in their own self-ratings of morality ($F(3, 90) = 9.95$ $p < .001$, $\eta^2 = .24$), but also confirmed a highly revealing significant interaction between the two factors ($F(3, 90) = 4.46$ $p < .001$, $\eta^2 = .13$).

Table 1 shows the frequencies per condition in terms of Batson et al.'s (2002) task assignment behavior categories, and the mean rated morality of the way they assigned the tasks by participants for each category. The abovementioned interaction reveals that the participants who dutifully flipped the coin (i.e., those participants in the task assignment behavior categories "Assigned other to positive task" and "Flip the coin, win and assign self to positive task") in the conditions where the experimenter had provided the "salient"-recommended moral norm (i.e., in conditions: Conflicting Norms and Moral Norm) rated the morality of their decision as more positive ($M = 8.39$) than did participants who adhered to the coin flip in conditions where the experimenters merely made random assignment salient with no "salient"-recommended moral norm (i.e., in conditions: Discretionary Norm and Obedient Cooperation, $M = 6.92$, $t(34.69) = 2.57$, $p < .05$, two tails).

Meanwhile, those participants in the conditions with no "salient" or explicit moral norm (i.e., Discretionary Norm) who did not flip the coin (i.e., those participants in the task assignment behavior category "Not flip the coin, assign self to positive task") rated the morality of their decisions as less negative ($M = 5.45$) than did the participants who did so in the conditions with a "salient"-recommended moral norm (i.e., Conflicting Norms and Moral Norm, $M = 3.46$, $t(23) = 1.93$, $p = .06$, two tails).

The participants' ratings of the morality of their own decisions were consistent with the hypothesis that the participants' recognized and were attempting to comply with the moral recommendations suggested by their experimenters' "salience" instructions and were actually contrary to the presumed influence of hypocrisy motivation

Evidence of hypocrisy by using the Batson, et al.'s (2002) definitions of hypocrisy. Examination of Table 1 reveals that the Conflicting Norms condition replicated Batson et al.'s (2002) evidence that they labeled hypocrisy: those who assigned the other to the positive-consequences task or flipped the coin, won, and assigned themselves to the positive-consequences task ($N = 11$) rated their assignation as more moral than those ($N = 17$) who did not flip or fiddled the coin flip and assigned themselves to the positive task ($M = 8.46$ vs. $M = 4.37$, $t(20.25) = 5.01$, $p < .001$, two tails). But, more important, whereas there were no significant differences in the ratings of those who assigned other to the positive-consequences task and those who flipped the coin and win, participants who fiddled the coin flip and assigned themselves to the positive task rated the morality of their action significantly higher than did those who did not flip the coin and assigned themselves to the positive task ($M = 6.00$ vs. $M = 2.75$, $t(15) = 2.52$, $p < .05$, two

[3] Following this criterion, 11 participants (5 in the Conflicting Norms condition, 3 in the Moral Norm condition, and 3 in the Discretionary Norm condition) who did not flip the coin were assigned to the "fiddle the coin flip, assign self to positive task".

tails). According to Batson et al. (2002, p. 336), this last comparison is particularly relevant for verifying the hypocrisy hypothesis.

This "hypocritical" pattern, however, was not observed in those conditions in which the participants were requested to follow a moral norm (Moral Norm) or an explicit instruction (Obedient Cooperation). In the Moral Norm condition, those who assigned the other to the positive task or flipped the coin and won ($N = 19$) rated the morality of their action as significantly more moral than the minority of participants ($N = 9$) who did not flip or fiddled the coin flip and assigned themselves to the positive task ($M = 8.31$ vs. $M = 4.75$, $t(9.32) = 3.55$, $p < .01$, two tails). Nevertheless, there were no significant differences between those who fiddled the coin flip and those who did not flip and assigned themselves to the positive task ($M = 4.17$ vs. $M = 5.33$, $t(7) = .53$, $p = .61$). In the Obedient Cooperation condition, there were no "dishonest" participants at all, and no significant differences between those who assigned the other to the positive task and those who flipped and won ($M = 7.11$ vs. 6.27, $t(18) = .56$, $p = .57$).

Finally, in the Discretionary Norm condition – where random assignment was salient but no experimenter recommended a different pattern – there were no significant differences between those who assigned the other to the positive task or flipped and won ($N = 8$), and those who did not flip or fiddled ($N = 14$) ($M = 7.16$ vs. 6.89, $t(19.93) = 1.19$, $p = .24$). No one-to-one comparison between task assignment categories was statistically significant in this condition (see Table 1).

Evidence of compliance

Our findings show that the experimenters' instructions described as intended to make a moral standard "salient" actually functioned to tell the participants what the experimenter believed to be "most moral" and that the experimenter wanted them to recognize and follow that as a recommended response. Participants' morality ratings of flipping the coin and their own decision were significantly different (see above paragraph on Participants' ratings) depending on the experimenter's instructions.

Of course, as good participants they would choose not to openly disagree with their experimenters' stated preferences. Thus the public compliance with the moral standard the experimenter promoted. A classic illustration of how authority engenders compliance is Milgram's (1974) Experiment 7: a significant portion of those participants who were not in the presence of the experimenter "administered lower shocks than were required and never informed the experimenter of the deviation (...) although these subjects acted in a way that clearly undermined the avowed purposes of the experiment, they found it easier to handle the conflict in this manner than to precipitate an open break with authority" (Milgram, 1974, p. 62).

An additional piece of information from the debriefing supports the otherwise well-documented desire not to offend the experimenter. Besides the above discussed ratings of morality, the post-experimental questionnaire asked participants which was the most morally right way of assigning the task. Interestingly, participants' open-format answers that were available to the experimenter showed an even clearer pattern of compliance with the experimenter's induction: in the condition in which the experimenter does not induce or order a particular decision (Discretionary Norm) only 8 (36%) participants chose random assignment as the most moral way of assigning the task. In the other three conditions a vast majority (80%, 69%, and 82%) chose this option (Chi Square (3, $N = 104$) = 14.43, $p = .002$, two tails). In other words, participants from the condition with no experimenter's induction were the only ones who made clearly independent judgments about their moral preferences.

Batson et al. (1997, 1999, 2002) published findings provide additional evidence supporting our data. For example, in those experiments where their "salient" moral standard consisted of random assignment by flipping the coin, the participants' subsequent reports of the "most moral" act simply repeated that (e.g., Batson et al., 1999, Study 2). However, in other experiments when the experimenters' "most moral" instructions made altruism "salient", the participants agreed that assigning the other to the desirable condition was most moral, not flipping a coin (e.g., Batson et al., 1999, Study 3). And finally, when the experimenter had not made either random assignment or assigning the other to the desirable condition "salient" neither of those were selected by the majority of participants as the "most moral" way to assign the conditions: 35% selected assigning the other, only 18% selected random assignment as most moral, and 32% stated that there was "no morally right way" to assign the conditions (Batson et al., 1999, Study 3, Low Standard Salience).

That leaves the question of why many of the participants' often elected, in private, to take the desirable condition for themselves – rather than doing what the experimenter recommended as most fair or moral? According to Batson et al., that was clearly a moral failure on the participants' part, intended to egoistically benefit themselves.[4] But this may not have been an egoistic act at all. In Batson et al.'s experiments, instructions had explicitly told participants that they were entitled to do as they wished and then provided them with the privacy in order to take the desirable condition for themselves. After all, if the experimenters had not expected the participants to take the desirable condition for themselves, why would they have then told the participants that they were going to deceive the other participants into believing that their "salient" moral standard of random assignment had been followed in assigning them to the relatively undesirable conditions? In effect, the experimenter's provision of privacy and planned deception provided the conflicted participants with a recommended solution: "Follow the norm of justified self-interest in the privacy I provided you and then, *just as I did*, adhere to the moral standard of random assignment in public."

In sum, our data show that the supposedly hypocritical students in Batson's research were merely responding appropriately as good cooperative participants. In this context (Conflicting Norms condition), for many of these participants labeled as hypocrites that meant following the experimenters' suggestions to behave in private according to the norm of justified self-interest, and then publicly support the experimenters' moral recommendations. When the experimenter instructions unambiguously induced (Moral Norm condition) or openly instructed (Obedient Cooperation condition) random assignment we found no "hypocrisy". On the other hand, when participants were free to choose the most legitimate norm, they openly adopted a justified self-interest option.

Conclusion

In conclusion, the results observed in our research suggest the need to re-examine the purported failures of moral integrity observed in the research by Batson and his colleagues. Our first condition (Replication Random Assignment), which replicated Batson's paradigm on hypocrisy, revealed participants' remarkable, and at times, creative compliance with two normatively supported

[4] Some readers could argue that participants' nonconsequential flipping of the coin is already a sign of moral feebleness or hypocrisy. Actually, participants' flipping of the coin might not only be related to their public endorsement of the morality of a random choice but also to a cognitive bias (Shafir, 1994; Tversky & Shafir, 1992). When people are forced into disjunctive situations, they seek for any available additional information that apparently has an effect on their choice but actually is not going to be implemental for the final decision: a substantial number of those participants who were clearly inclined to keep the positive task for themselves, irrespective of the flipping outcome, might have been willing to postpone their decision until flipping the coin, out of reluctance to think through disjunctions under a seeming uncertainty.

"demands" communicated by the experimenters instructions. When two contradictory demands were communicated – behavioral support for justified self-interest and verbal recommendations to follow the "most moral" standard of random assignment – many of the participants mimicked what the experimenter had modeled for them, and elected the desirable condition for themselves in private but then gave lip-service to the experimenters' suggestion of the "fairest" norm in the public ratings.[5]

Most of the remaining participants in the other three conditions were not confronted with contradictory normative expectations, but both their behavior and public ratings revealed that they recognized the expectations associated with the norm of justified self-interest and the experimenters' recommendation concerning the fairest standard of behavior. When either norm was invoked by the experimenters, the participants' behavior did not significantly deviate from that norm: as cooperative participants that meant either taking the desired outcome for themselves or following random assignment. In addition, that cooperation extended to their publicly adopting the moral standard recommended by the experimenter. That appeared in measurably lower ratings of the morality of their own behavior among those who deviated from the "most moral" standard recommended by the experimenter. That, of course, was not consistent with the assumption of hypocrisy motivation.

We do not mean to suggest that people are never hypocritical. To the contrary, the potential for hypocrisy is endemic in social life and has taken on societal significance in recent years as one institution after another – from corporate America to the Catholic Church – has been exposed preaching one thing and behaving in a quite different manner. Hypocrisy undermines trust and breeds cynicism, so people are understandably vigilant regarding its occurrence and hold disdainful attitudes towards those who practice it. Because hypocrisy is socially dysfunctional and hypocrites are derogated – that is, it is maladaptive both for social systems and for individual actors – it may constitute an aspect of human psychology that is tightly regulated and correspondingly infrequent.

Given these constraints, hypocrisy is likely to be confined to instances in which the expected gain for such behavior exceeds a critical threshold (cf. Feather, 1995) or when the actor experiences a breakdown in self-regulatory mechanisms (e.g., under cognitive load, when attention is focused away from the self, when the actor is disinhibited or in the grips of an irrepressible impulse) (cf. Baumeister, Schmeichel, & Vohs, 2007; Keltner, Gruenfeld, & Anderson, 2003; Metcalfe & Mischel, 1999; Wicklund & Frey, 1980). It would be a stretch to claim that the paradigm developed by Batson and his colleagues fulfills either of these requirements. Moral hypocrisy certainly exists, though, and we encourage further research that not only provides unequivocal evidence of its occurrence, but also grounds this unwanted feature of human psychology in the context of clearly delineated parameters.

A major contribution of the research reported here involves the re-introduction of the important distinction between legitimate and supererogatory moral norms. Recognizing that distinction should lead future investigators to a more thorough and hence valid consideration of what constitutes moral and immoral behavior for actors and bystanders in a given situation. Expanding the familiar dichotomous alternatives of moral-immoral, egoism-altruism to include an actor's legitimate – even though not the most laudatory – options opens the way to a more accurate view of an individual's motives when confronted with moral dilemmas inherent in an overwhelmingly complex world.

Acknowledgments

This research was partially supported by Grant PSI2008-04849 funded by the Spanish Government (MICINN).

References

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice. Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72*, 1335–1348.

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., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77*, 525–537.

Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 516–539). New York: Guilford Press.

Feather, N. (1995). Values, valences, and choice. The influence of values on the perceived attractiveness and choice of alternatives. *Journal of Personality and Social Psychology, 68*, 1135–1151.

Heyd, D. (2006). Supererogation. In *Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/entries/supererogation/> Retrieved 24.07.08.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 116*, 265–284.

Lerner, M. J. (1971). Justified self-interest: A replication and extension. *Journal of Human Relations, 19*, 550–559.

Lerner, M. J., & Lichtman, R. R. (1968). Effects of perceived norms on attitudes and altruistic behavior toward a dependent other. *Journal of Personality and Social Psychology, 9*, 226–232.

Lerner, M. J., Miller, D. T., & Holmes, J. G. (1976). Deserving and the emergence of forms of justice. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 9, pp. 134–160). New York: Academic Press.

Metcalfe, J., & Mischel, W. (1999). A hot/cool system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*, 3–19.

Milgram, S. (1974). *Obedience to authority: An experimental view*. London: Tavistock.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition, 50*, 403–420.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*, 305–309.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science, 18*, 689–690.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology, 44*, 1334–1338.

Wicklund, R. A., & Frey, D. (1980). Self-awareness theory: When the self makes a difference. In D. M. Wegner & R. R. Vallacher (Eds.), *The self in social psychology* (pp. 31–54). New York: Oxford University Press.

[5] Valdesolo and DeSteno (2008) reported that when participants were asked whether, in Batson's dilemma, a choice based on self-interest was moral or immoral, compared to a random choice, all of them identified self-interest as immoral (see also Batson et al., 1997, Study 1). Recognizing that they had failed to provide their participants with the important opportunity to elect "legitimate", we corrected that in a subsequent study. Thirty participants judged the decision of an anonymous student who assigned him or herself to the positive-consequences task in Batson's dilemma (see Batson et al., 1999; Experiment 3, Low Standard Salience Condition). The available ratings were "illegitimate", "legitimate" and "most legitimate decision imaginable". All participants (100%) rated the student's decision as "legitimate". The participants' explanations of their choices revealed that they discriminated between legitimate and supererogatory actions: 88% of them emphasized that the experiment's rules entitled the student to keep the positive task, and 28% added that by choosing the positive task the student did not hurt the other student involved in the experiment.