Preregistered Replication of "Moral Hypocrisy: Social Groups and the Flexibility of Virtue"

[redacted]

Abstract

The tendency for people to consider themselves morally good while behaving selfishly is known as “moral hypocrisy”. Influential work by Valdesolo & DeSteno (2007) found evidence for intergroup moral hypocrisy, such that people are more forgiving of transgressions when they are committed by an in-group member than an out-group member. We propose a direct and conceptual replication of Valdesolo and DeSteno’s work on moral hypocrisy and group membership. We plan to directly replicate their original study using minimal groups, as well as conceptually replicate their work using natural groups (i.e., political party affiliation). Our study will have implications for understanding moral hypocrisy, intergroup bias, and partisanship.

Keywords: morality, groups, identity, minimal groups, partisanship

Preregistered Replication of "Moral Hypocrisy:

Social Groups and the Flexibility of Virtue”

Although the average person holds themselves to a moral code, the average person is also able to commit immoral acts. These acts may range from the mundane, such as cutting in line, to the extreme, such as shocking someone to death (Milgram, 1963). Regardless, people continue to consider themselves moral beings even after committing immoral acts -- a phenomenon termed Moral Hypocrisy (Batson et al., 1997, Batson et al., 2002). Moral hypocrisy may stem from the cognitive need to justify our own immoral actions and place them in accordance with our perceptions of our own identity as moral beings (Shalvi et al., 2011; Shalvi et al., 2015). Critically, however, both one’s moral sense and identity are heavily influenced by the social groups they belong to (Graham, Haidt & Nosek, 2009). Social groups exert a powerful influence, such that people are likely to favor their ingroup members and derogate their outgroup members (Tajfel & Turner, 1979; Balliet, Wu & De Drue 2014; Leach et al. 2003; Rathje et al., 2021). Therefore, in the current research we ask the question: does moral hypocrisy extend beyond the self?

Influential work in this area has found that group identity shapes moral hypocrisy (Valdesolo & DeSteno, 2007). Using an elegantly simple study design, the researchers found that the same immoral action (assigning an easier task to oneself and a more onerous task to someone else) was judged to be more fair when the participant themselves, or a member of the participant’s ingroup, was the perpetrator. In other words, the morally hypocritical benefits that people give themselves extend to one’s in-group members, but not to one’s out-group members.

We seek to replicate Valdesolo and DeSteno's findings to determine if they generalize to a new sample over a decade later. The implications of these findings are far-reaching; this work suggests that moral convictions, which have been theorized to be inflexible and universal (Skitka, 2010; Van Bavel et al., 2012), are susceptible to in-group bias. Moral hypocrisy has strong social consequences for individuals, such that those who are viewed as hypocritical are found to be deserving of more punishment for a transgression compared to non-hypocrites (Effron et al., 2018, Barden, Rucker & Petty, 2005). Therefore, the extension of moral hypocrisy to social groups has important implications for intergroup conflict including political polarization (Finkel et al., 2020).

In our replication, we will improve upon Valdesolo & DeSteno's methodology in three main ways: (1) increasing the sample size and statistical power, (2) adding new explanatory analyses, and (3) extending the finding to real world groups to evaluate external validity. First, the original paper has a relatively small sample size. The total N of the study was 76, split into 4 conditions (providing 19 participants in each cell). The reported effect size is $d = 1.11$, which is very large for a social psychological study where the average effect size is closer to $d = .4$ (Richard et al., 2003).Using this estimate, we will increase our sample size to achieve 95% power at an alpha level of .01 estimating that the effect size is $d = .4$, the average effect size for social psychology studies.

Secondly, we will examine the moderating effect of strength of collective identification on intergroup moral hypocrisy. Prior work has suggested that the strength of one's identification with their ingroup is associated with increased perceived ingroup homogeneity and increased outgroup derogation (Hornsey, 2008; Leach et al., 2003; Branscombe et al., 1999). To examine

how individual differences in level of collective identification affect reported intergroup moral hypocrisy, we will conduct an exploratory multiple regression analysis adjusting for collective identification.

Third, the original procedure used a minimal groups design (Tajfel et al., 1971). While a minimal groups design offers an excellent and well controlled test of the moral hypocrisy effect, it is unclear if intergroup moral hypocrisy would generalize to real world groups, where moral hypocrisy *appears* to be quite prevalent (Wolsky, 2020; Cottle, 2021). We plan to replicate the effect in both minimal groups and natural groups to increase external validity. In everyday life, moral conflict (including moral hypocrisy) is most likely to occur between groups that have historical and/or sociological origins such as religion (Ginges et al. 2007) or political affiliation (Brady et al. 2020; Finkel et al. 2020). As such, we will conduct a novel Study 2 in which participants follow the same procedure as above, but will be separated based on natural groups consisting of their political party identification (i.e., Democrats or Republicans) instead of minimal groups. This will determine if the main findings generalize to real world groups. Otherwise, the procedure laid out in the original paper is comprehensive, and we propose following the original procedure almost exactly in order to conduct a direct replication.

We hypothesize that (1) people's evaluations of their own fairness will be significantly higher compared to their evaluations of others' fairness after committing the same moral transgression. We further hypothesize that (2) people will evaluate their ingroup members as acting significantly more fairly than outgroup members after committing the same moral transgressions. We hypothesize that (3) this effect will be present when ingroups and outgroups are defined by both minimal groups and natural groups (political party identification). Finally,

we hypothesize that (4) the strength of collective identification will mediate moral hypocrisy, such that people who are strongly identified with their group will rate their ingroup member's actions as more fair and their outgroup member's actions as less fair.

Methods

Participants

To increase the statistical power from the original paper, we plan to increase the total sample size as well as the cell sample sizes for each condition in both Study 1 and Study 2 (see Brandt et al., 2014). To find out how many participants, we conducted a power analysis in G*Power assuming the average medium effect size common in psychology, $d = .4$ (Richard et al., 2003). According to a power analysis for a 4-group one-way omnibus ANOVA with a Cohen's $f = .2$ (equivalent to $d = .4$) and an alpha of .01, we would need 576 total participants split into 4 groups to achieve 95% power. Therefore, each cell in both Study 1 and Study 2 will have n = 144. These parameters have the benefit of replicating the findings from the original paper using the most stringent significance criteria.

We also ran a power analysis for equivalence testing (specifically, TOST; two one-sided t tests) using the TOSTER R package function 'powerTOSTone.' We tested whether we would have the power to reject the presence of effects of $d > 0.2$. According to this power analysis, with an alpha of .01 and the proposed sample size of 576, we would have .97 power. We plan to achieve a sample size of 576 using an online survey platform such as Prolific. We will recruit a politically representative sample for both Study 1 and Study 2 (a ratio of liberals and conservatives that approximates the U.S. adult population).

Procedure

Study 1

Before the experiment begins, participants will fill out a short survey that includes questions about their political affiliation, the overestimator/underestimator task with false feedback (Tajfel et al., 1972), and distractor questions. At the beginning of the experiment, participants in all conditions read instructions stating that researchers are interested in performance on two different tasks. Task 1 (the "green" task in the original paper) is a simple 10 minute task, and Task 2 (the "red" task in the original paper) is a complex 45 minute task. Participants will also be told that, in order to keep the researchers blind to the condition of each participant, the researchers are using a newly developed assignment procedure for which a subset of participants will be selected to choose which task to assign to themselves and another participant. Participants will be told that those chosen to make assignments can either assign tasks randomly, using a computer randomizer, or they can select one task for themselves, leaving a future participant to complete the other, unselected task. Therefore, when participants select the less onerous task for themselves even though a fair selection option is available (i.e. the randomizer), they commit a moral transgression.

In Condition 1, participants themselves will be instructed to select which task they would like to complete. In the original paper, 17 participants out of a total 19 participants in Condition 1 assigned themselves the less onerous task. The two participants who chose to act altruistically (one using the computer randomizer and one choosing the worse task for themselves) were excluded from analyses. We will also exclude altruistic participants from all analyses.

Once task selection is completed, participants will answer questions about the "experimenter blind" selection procedure. Embedded in the questionnaire, there will be a question about how fairly the participant thought they acted when allocating the tasks. This is the dependent measure of interest, and is answered on a 7-point Likert scale (from 1 = "*extremely unfairly*" to 7 = "*extremely fairly*").

In Condition 2, participants follow the same procedure as Condition 1, except that participants are told they will observe another participant use the newly developed assignment procedure to allocate the "red" task and the "green" task for future participants. Participants are told that the allocator has the choice to use a randomizer, or to assign themselves to the Participants then watch a computer confederate behave selfishly, assigning themselves the "green" task and assigning a future participant the "red" task. No other information will be given about this "other" participant.

Condition 3 will follow the same procedure as Condition 2, except that participants will be told that the participant they observe assigning tasks is part of the same minimal group that the participant is. Participants will read the following information: "You will now watch another participant assign conditions to themselves and a future participant. You've been randomly assigned to see the following information about this participant from their survey: They are a [Overestimator/Underestimator]." This will be done to ensure they are judging an in-group member. Condition 4 will follow the same procedure as Condition 3, except that participants will be told that the observed participant is a member of their minimal out-group. As an additional measure in our replication, participants in Condition 3 and Condition 4 will also complete a 3-item measure of collective identification common in social identity literature with minimal

groups and natural groups (Ashmore et al., 2004; Van Bavel & Cunningham, 2012). The three questions are "I value being a member of the Overestimator (Underestimator) group," "I am proud to be a member of the Overestimator (Underestimator) group," and "Belonging to the Overestimator (Underestimator) group is an important part of my identity." The questions will be presented in random order to each participant.

Study 2

The procedure for Study 2 is identical to Study 1, except that instead of using minimal groups to establish participants' in-group and out-group membership, we will ask participants to report their political party affiliation, as well as how strongly they identify with their party. Prolific allows researchers to collect an equal number of Democratic and Republican participants, eliminating the Democratic bias in many online survey groups (Huff & Tingley, 2015). Furthermore, each condition will have equal numbers of Democrat and Republican participants. Participants in Condition 1 will follow the exact same Condition 1 procedure laid out in Study 1. Condition 2 will follow exactly the same procedure as in Study 1, where they will watch another participant behave selfishly and learn nothing about that participants' identity. In Condition 3, participants will watch someone from their political in-group behave selfishly (e.g., a participant who identifies as a Republican will be told they are watching a Republican). To inform participants of the target's political in-group status, the participant will read the instructions as follows "You will now watch another participant assign conditions to themselves and a future participant. You have been randomly assigned to see the following information about this participant from their survey: They are a [Democrat/Republican]." In Condition 4, participants will watch someone from their political outgroup behave selfishly (e.g., a participant

who identifies as a Republican will be told they are watching a Democrat). Revealing the target's political party affiliation will be similar to Condition 3. Participants in Condition 3 and Condition 4 will complete the same 3-item measure of collective identification as in Study 1.

Analysis Plan

The original manuscript analyzed their findings using a planned contrast, where Condition 1 and Condition 3 had contrast weights of 1 and Condition 2 and Condition 4 had contrast weights of -1. In our analyses, both Study 1 and Study 2 will undergo a one-way omnibus ANOVA before planned contrasts. We will then conduct a planned contrast as outlined above to exactly replicate the original analysis. We will also conduct an exploratory multiple regression analysis to test the effect of collective identification on fairness ratings for both studies' Conditions 3 and 4. Specifically, we will regress fairness ratings on collective identification, adjusting for group status of the confederate.

As an exploratory analysis, we will compare the two studies directly using a 2 (Group Type) X 4 (Condition) ANOVA. We expect to see that the moral hypocrisy effect is stronger among the natural group participants than the minimal group participants. Regarding concerns about floor or ceiling effects in statistical analyses, there is no reason in the original paper to suspect that floor or ceiling effects will occur.

There will be data quality checks in the study. The first will come just after the consent form, and will ask participants to label a simple image to detect bots. The second data quality check will come after the mock survey and will be an attention check designed to look like a long question. At the end of the question, participants will be instructed to click on the NYU

logo in the top of the screen, and not use the radio buttons. Any participant that fails either of the attention checks will be excluded from the data.

Expected Results

Study 1

We predict that our results will largely replicate the results of the original Valdesolo & Desteno (2007) paper in Study 1. Participants in both Condition 1 (Self) and Condition 3 (Minimal Ingroup Other) will rate their actions as significantly fairer than participants in Condition 2 (Unaffiliated Other) and Condition 4 (Minimal Outgroup Other) (see Figure 1 in the supplemental materials for expected results). These results would provide support for the existence of moral hypocrisy in the self, as well as the extension of moral hypocrisy to in-group members.

We also predict that collective identification will moderate ingroup hypocrisy, such that fairness ratings of one's ingroup will increase as collective identification increases in Condition 3, and fairness ratings of one's outgroup member will decrease as collective identification increases in Condition 4.

Study 2

We predict that we will also conceptually replicate the results of the original Valdesolo and Desteno (2007) paper in Study 2 using a new, natural groups paradigm. Participants in both Condition 1 (Self) and Condition 3 (Political Ingroup Other) will rate their actions as significantly more fair than participants in Condition 2 (Unaffiliated Other) and Condition 4 (Political Outgroup Other). In conclusion, the findings from both studies outlined in this proposal would provide empirical support that the moral hypocrisy effect is robust. The proposed studies

increase statistical power and external validity, as well reproducibility for a seminal finding in the field of social psychology.

References

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: articulation and significance of multidimensionality. *Psychological bulletin*, 130(1), 80–114.

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, 140(6), 1556.

Barden, J., Rucker, D. D., & Petty, R. E. (2005). "Saying one thing and doing another": examining the impact of event order on hypocrisy judgments of others. *Personality & social psychology bulletin*, 31(11), 1463–1474.

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*(6), 1335.

Batson, C. D., Thompson, E. R., & Chen, H. (2002). Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology*, *83*(2), 330.

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science, 15*(4), 978-1010.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication?. *Journal of Experimental Social Psychology, 50*, 217-224.

Cottle, M. (2021, May 4). *Who Cares About Hypocrisy?* The New York Times. https://www.nytimes.com/2021/05/04/opinion/republicans-biden-hypocrisy.html

Effron, D., Markus, H., Jackman, L., Muramoto, Y., Muluk, H. (2018). Hypocrisy and culture: Failing to practice what you preach receives harsher interpersonal reactions in independent (vs. interdependent) cultures. *Journal of Experimental Social Psychology, 76,* 371-384.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., ... & Druckman, J. N. (2020). Political sectarianism in America. *Science, 370*(6516), 533-536.

Ginges, J., Atran, S., & Medin, D. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences, 104*, 7357-7360.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5), 1029.

Hornsey, M. J. (2008). Social identity theory and self‐categorization theory: A historical review. *Social and personality psychology compass*, 2(1), 204-222.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3), 2053168015604648.

Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of personality and social psychology*, 84(5), 932.

Milgram, S. (1963). Behavioral study of obedience. The Journal of abnormal and social psychology, 67(4), 371.

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26).

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of general psychology, 7*(4), 331-363.

Shalvi, S., Dana, J., Handgraaf, M., & De Dreu, C.K.W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes,* 115(2), 181-190.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science*, 24(2), 125–130.

Skitka, L. J. (2010). The psychology of moral conviction. Social and Personality Psychology *Compass, 4*(4), 267-281.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European journal of social psychology, 1*(2), 149-178.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: social groups and the flexibility of virtue. *Psychological Science*.

Van Bavel, J. J., & Cunningham, W. A. (2012). A Social Identity Approach to Person Memory: Group Membership, Collective Identification, and Social Role Shape Attention and Memory. *Personality and Social Psychology Bulletin*, 38(12), 1566–1578.

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PloS one, 7*(11), e48693.

Wolsky, A. (2020). Scandal, Hypocrisy, and Resignation: How Partisanship Shapes Evaluations of Politicians' Transgressions. *Journal of Experimental Political Science*, 1-14.

Figures

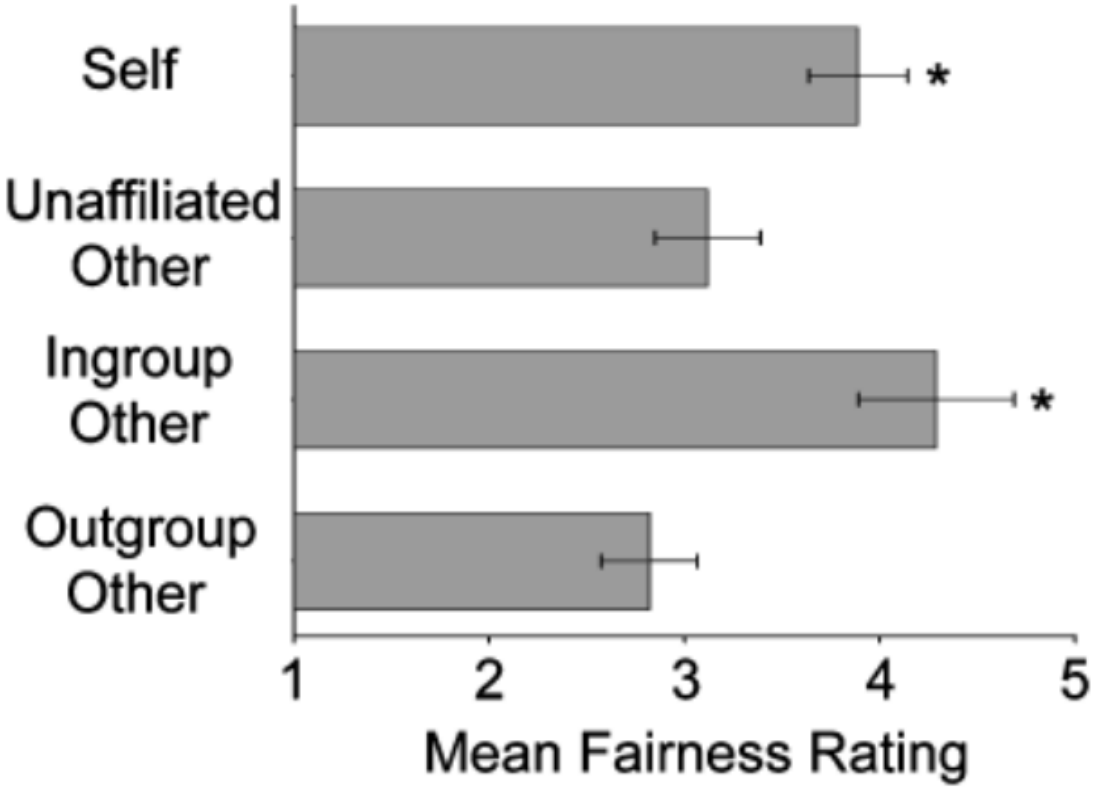

Figure 1: Original image from Valdesolo & DeSteno (2007) depicting mean fairness ratings for the self (Condition 1), an unaffiliated other (Condition 2), an ingroup other (Condition 3) and an outgroup other (Condition 4).