

# Social Psychology

# 3/14

[www.hogrefe.com/journals/sp](http://www.hogrefe.com/journals/sp)

**Editor-in-Chief**

Christian Unkelbach

**Associate Editors**

Julia Becker · Malte Friese

Michael Häfner · Eva Jonas

Markus Kimmelmeier · Ulrich Kühnen

Alison Ledgerwood · Michaela Wänke

**Special Issue**

Replications of Important  
Results in Social Psychology

**Guest Editors**

Brian A. Nosek

Daniël Lakens

HOGREFE



---

# Contents

<b>Editorial</b>	Registered Reports: A Method to Increase the Credibility of Published Results <i>Brian A. Nosek and Daniël Lakens</i>	137
<b>Replications</b>	Investigating Variation in Replicability: A “Many Labs” Replication Project <i>Richard A. Klein, Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams Jr., Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, Mark J. Brandt, Beach Brooks, Claudia Chloe Brumbaugh, Zeynep Cemalcilar, Jesse Chandler, Winnee Cheong, William E. Davis, Thierry Devos, Matthew Eisner, Natalia Frankowska, David Furrow, Elisa Maria Galliani, Fred Hasselman, Joshua A. Hicks, James F. Hovermale, S. Jane Hunt, Jeffrey R. Huntsinger, Hans IJzerman, Melissa-Sue John, Jennifer A. Joy-Gaba, Heather Barry Kappes, Lacy E. Krueger, Jaime Kurtz, Carmel A. Levitan, Robyn K. Mallett, Wendy L. Morris, Anthony J. Nelson, Jason A. Nier, Grant Packard, Ronaldo Pilati, Abraham M. Rutchick, Kathleen Schmidt, Jeanine L. Skorinko, Robert Smith, Troy G. Steiner, Justin Storbeck, Lyn M. Van Swol, Donna Thompson, A. E. van 't Veer, Leigh Ann Vaughn, Marek Vranka, Aaron L. Wichman, Julie A. Woodzicka, and Brian A. Nosek</i>	142
	Forming Impressions of Personality: A Replication and Review of Asch's (1946) Evidence for a Primacy-of-Warmth Effect in Impression Formation <i>Sanne Nauts, Oliver Langner, Inge Huijsmans, Roos Vonk, and Daniël H. J. Wigboldus</i>	153
	Revisiting Schachter's Research on Rejection, Deviance, and Communication (1951) <i>Eric D. Wesselmann, Kipling D. Williams, John B. Pryor, Fredrick A. Eichler, Devin M. Gill, and John D. Hogue</i>	164
	Revisiting the Romeo and Juliet Effect (Driscoll, Davis, & Lipetz, 1972): Reexamining the Links Between Social Network Opinions and Romantic Relationship Outcomes <i>H. Colleen Sinclair, Kristina B. Hood, and Brittany L. Wright</i>	170
	Breakthrough or One-Hit Wonder? Three Attempts to Replicate Single-Exposure Musical Conditioning Effects on Choice Behavior (Gorn, 1982) <i>Ivar Vermeulen, Anika Batenburg, Camiel J. Beukeboom, and Tim Smits</i>	179
	What Does It Take to Activate Stereotypes? Simple Primes Don't Seem Enough: A Replication of Stereotype Activation (Banaji & Hardin, 1996; Blair & Banaji, 1996) <i>Florian Müller and Klaus Rothermund</i>	187
	A Replication Attempt of Stereotype Susceptibility (Shih, Pittinsky, & Ambady, 1999): Identity Salience and Shifts in Quantitative Performance <i>Carolyn E. Gibson, Joy Losee, and Christine Vitiello</i>	194

---

# Contents

A Secondary Replication Attempt of Stereotype Susceptibility (Shih, Pittinsky, & Ambady, 1999) <i>Alice Moon and Scott S. Roeder</i>	199
Sex Differences in Distress From Infidelity in Early Adulthood and in Later Life: A Replication and Meta-Analysis of Shackelford et al. (2004) <i>Hans IJzerman, Irene Blanken, Mark J. Brandt, J. M. Oerlemans, Marloes M. W. Van den Hoogenhof, Stephanie J. M. Franken, and Mathe W. G. Oerlemans</i>	202
Does Cleanliness Influence Moral Judgments? A Direct Replication of Schnall, Benton, and Harvey (2008) <i>David J. Johnson, Felix Cheung, and M. Brent Donnellan</i>	209
Replication of “Experiencing Physical Warmth Promotes Interpersonal Warmth” by Williams and Bargh (2008) <i>Dermot Lynott, Katherine S. Corker, Jessica Wortman, Louise Connell, M. Brent Donnellan, Richard E. Lucas, and Kerry O’Brien</i>	216
Replication of Experiments Evaluating Impact of Psychological Distance on Moral Judgment: (Eyal, Liberman & Trope, 2008; Gong & Medin, 2012) <i>Iris L. Žeželj and Biljana R. Jokić</i>	223
Three Attempts to Replicate the Moral Licensing Effect <i>Irene Blanken, Niels van de Ven, Marcel Zeelenberg, and Marijn H.C. Meijers</i>	232
Replication of the Superstition and Performance Study by Damisch, Stoberock, and Mussweiler (2010) <i>Robert J. Calin-Jageman and Tracy L. Caldwell</i>	239
Does Recalling Moral Behavior Change the Perception of Brightness? A Replication and Meta-Analysis of Banerjee, Chatterjee and Sinha (2012) <i>Mark J. Brandt, Hans IJzerman, and Irene Blanken</i>	246

# Social Psychology

Your article has appeared in a journal published by Hogrefe Publishing. This e-offprint is provided exclusively for the personal use of the authors. It may not be posted on a personal or institutional website or to an institutional or disciplinary repository.

If you wish to post the article to your personal or institutional website or to archive it in an institutional or disciplinary repository, please use either a pre-print or a post-print of your manuscript in accordance with the publication release for your article and our “Online Rights for Journal Articles” ([www.hogrefe.com/journals](http://www.hogrefe.com/journals)).

# Three Attempts to Replicate the Moral Licensing Effect

Irene Blanken,<sup>1</sup> Niels van de Ven,<sup>1</sup> Marcel Zeelenberg,<sup>1</sup> and Marijn H.C. Meijers<sup>2</sup>

<sup>1</sup>Tilburg University, The Netherlands, <sup>2</sup>University of Amsterdam, The Netherlands

**Abstract.** The present work includes three attempts to replicate the moral licensing effect by Sachdeva, Iliev, and Medin (2009). The original authors found that writing about positive traits led to lower donations to charity and decreased cooperative behavior. The first two replication attempts (student samples, 95% power based on the initial findings,  $N_{\text{Study1}} = 105$ ,  $N_{\text{Study2}} = 150$ ), did not confirm the original results. The third replication attempt (MTurk sample, 95% power based on a meta-analysis on self-licensing,  $N = 940$ ) also did not confirm the moral licensing effect. We conclude that (1) there is as of yet no strong support for the moral self-regulation framework proposed in Sachdeva et al. (2009) (2) the manipulation used is unlikely to induce moral licensing, and (3) studies on moral licensing should use a neutral control condition.

**Keywords:** moral licensing, moral cleansing, self-regulation, replication

People like to present themselves as good people, both to themselves and to others, to maintain a positive self-image and to feel like a moral person (Aronson, Cohen, & Nail, 1999; Schlenker, 1980; Steele, 1988). Furthermore, central theories of human behavior highlight humans' desire for cognitive consistency in their thoughts, feelings, and behavior (Festinger, 1957; Heider, 1946). Intriguing research on *moral licensing* qualifies this desire for consistency by suggesting that individuals who behave in a morally laudable way, later feel more justified to perform a morally questionable action (Merritt, Effron, & Monin, 2010; Miller & Effron, 2010). Moral licensing is found to lead to a broad spectrum of undesirable behaviors. For example, after (reminders of) prior moral or socially desirable behavior people displayed more prejudiced attitudes (Effron, Cameron, & Monin, 2009; Monin & Miller, 2001), cheated more (Jordan, Mullen, & Murningham, 2011; Mazar & Zhong, 2010), displayed a preference for hedonic over utilitarian products (Khan & Dhar, 2006), and indulged more in highly palatable foods (Mukhopadhyay, Sengupta, & Ramanathan, 2008).

An important contribution to the literature on moral licensing examines how writing about one's own positive or negative traits can influence donations to charity and cooperative behavior in a commons dilemma (Sachdeva, Iliev, & Medin, 2009). In just 4 years since publication, this paper has been cited 129 times (Google Scholar, November 27, 2013). Based on their findings, the authors argued that this moral licensing effect can best be interpreted as part of a larger moral self-regulation framework where internal

balancing of moral self-worth and the costs associated with prosocial behavior determine whether one will display (im)moral behavior. When the moral image of oneself is established, an immoral action is allowed without the fear of losing that moral image (moral licensing). However, when one appears immoral to others, positive actions are needed to restore the moral image (moral cleansing). The studies of Sachdeva et al. (2009) comparing licensing with neutral control conditions show medium-sized effect sizes ( $d = 0.62$  ([CL<sub>95</sub>] -0.11 to 1.35) for Study 1 and  $d = 0.59$  ([CL<sub>95</sub>] -0.12 to 1.30) for Study 3).<sup>1</sup> However, note that because of the small sample sizes ( $N = 14$  to 17 per condition), the obtained effects have large variances, implying that the true effect sizes could range from very small to very large.

There are no published direct replication attempts of the methodologies of Sachdeva et al.'s (2009) studies. Conway and Peetz (2012) conducted a study that was similar to Sachdeva et al.'s Study 1. However, this was not a direct replication because they adapted the procedure and added extra manipulations. We sought to replicate the studies by Sachdeva et al. to obtain additional insight in the complete moral self-regulation framework by testing for both moral licensing and moral cleansing effects contrasted to a neutral control condition.

We conducted high-powered replications of Sachdeva et al.'s (2009) Study 1 and Study 3 in Dutch student samples with 95% statistical power based on the effect size of the original studies. We did a third study with a US sample via Amazon's MTurk with 95% power based on

<sup>1</sup> Note that the overall differences between the three conditions (moral licensing, moral cleansing, and the neutral control condition) of Sachdeva et al.'s Study 1 and Study 3 were significant. For Study 1, no statistics on post hoc comparisons were reported. When calculating the Cohen's  $d$  effect sizes comparing the moral licensing with the neutral control conditions, we found that for both studies, the confidence intervals included zero, indicating marginally significant moral licensing effects.

the effect size that we obtained in our meta-analysis on self-licensing ( $d = 0.26$ ; Blanken, Van de Ven, & Zeelenberg, 2014). This study examined both dependent variables of original Studies 1 and 3 in a counterbalanced order. For all studies, we report how we determined our sample sizes, all data exclusions, all manipulations, and all measures.

## Study 1 – Replication of Sachdeva et al.'s (2009) Study 1

### Participants

Using G\*Power (Faul, Erdfelder, Buchner, & Lang, 2009) we calculated that at least 63 participants were needed to achieve 95% power for the effect size of Sachdeva et al.'s Study 1 (2009;  $N = 46$ ). We planned to collect data for one full week, and our sample consisted of 106 undergraduate students who participated for course credit. One was removed because this participant indicated a willingness to donate €100 to charity, more than 32 standard deviations from the mean donation response. The remaining 105 (25 males, 78 females, 2 unknown,  $M_{\text{age}} = 19.58$ ) participants included native Dutch students (83.3%), non-native Dutch students (9.5%), and foreign students (4.8%). Participants were randomly assigned to either the positive trait ( $N = 35$ ), negative trait ( $N = 34$ ), or neutral control condition ( $N = 36$ ).

### Materials and Procedure

Participants completed the study as the first of a series of experiments behind separate desks in the laboratory. The experimenter in the laboratory was blind to condition. Prior to the experiment, participants provided their informed consent. The experimenter guided participants to their desks and instructed them to complete the paper-and-pencil questionnaire.

We obtained the original paper-and-pencil questionnaire from Sachdeva et al. (2009) and translated these materials into Dutch (for all materials see supplements section). The cover story indicated that the study was about handwriting styles. Depending on the assigned condition, participants were exposed to either nine positive trait words, nine negative trait words, or nine neutral words and were asked to copy each word four times and think about each word for 5–10 s. Next, participants were asked to write a short story about themselves including the words they just copied.

After this manipulation, participants responded to some filler items. Subsequently, the main dependent variable was presented. Participants read that the laboratory, in an effort to increase social responsibility, asked all participants whether they would like to contribute to a worthy cause.

If they would like to do so, they could pledge to make a small donation to any good cause of their choice. They were told that they would be reminded of their choice at a later time via a confirmation e-mail from the experimenter. Participants could select to which cause(s) they would like to donate (cancer research, animal rights, ending world hunger, environmental preservation, human rights, veteran's affairs, or other) and how much they would be willing to donate (from €0 up to €10 or another specified amount). Finally, participants completed seven self-presentation items from the Self-Monitoring scale (Lennox & Wolfe, 1984) and a set of demographic measures.

### Known Differences From Original Study

The only known difference between our replication and the original Studies 1 and 2 of Sachdeva et al. (2009) was that we ran this study in a laboratory at a Dutch university, while the original study was conducted in a laboratory at a USA university. When participants were asked to write about the positive trait, neutral, or negative trait words, we used the exact instruction of the original Study 2, which explicitly stated that participants should use the nine given words to *write a story about themselves*. This was not done in Sachdeva et al.'s Study 1, although it was intended that participants would do so. As such, for this replication, we combined the best of Sachdeva et al.'s Study 1 (including a control condition) and Study 2 (the manipulation with the clearest instruction).

### Results

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analysis to test the effect of writing about one's own positive traits, negative traits, or neutral words on donation amount. Table 1 contains the mean responses per condition and statistical tests. There were no significant differences between the moral identity conditions on donation amount.<sup>2</sup> The results of an additional regression model including gender, age, and ethnicity indicated that none of these factors significantly predicted donation amount (all  $ps \geq .321$ ). A reviewer suggested that self-monitoring might moderate the observed effects. It did not,  $p = .086$ . Analysis details for all studies are available in the supplements.

### Exploratory Analysis

When reading the recalled stories, we noticed that 55.7% of the participants violated the instructions by not writing about themselves or by using the words in a negating way (for instance, "Alyssa is a generally friendly person

<sup>2</sup> A nonparametric independent-samples Kruskal-Wallis test (which controls for the skewness of the data), also found a nonsignificant effect,  $H(2) = 0.36, p = .837$ .

with a caring and compassionate disposition” or “I am neither a very caring nor compassionate individual”). When we only used a post hoc selection of those that wrote about their own positive traits ( $N = 28$ ) and compared it to the neutral control condition, there was still no difference on donation amount ( $p = .756$ ).

## Study 2 – Replication of Sachdeva et al.'s (2009) Study 3

### Participants

Using G\*Power (Faul et al., 2009) we calculated that we should include at least 96 participants in our study to achieve 95% power for the effect size that Sachdeva et al. (2009) obtained in their Study 3 (the original used  $N = 46$ ). We planned to collect data for one full week, and our sample consisted of 150 undergraduate students who participated for course credit (27 males, 122 females, 1 unknown,  $M_{age} = 20.34$ ) and included native Dutch students (87.3%), non-native Dutch students (7.3%), and foreign students (4.7%). All participants were randomly assigned to either the positive trait condition ( $N = 49$ ), the negative trait condition ( $N = 52$ ), or the neutral control condition ( $N = 49$ ).

### Materials and Procedure

Participants first provided informed consent, and then completed the study as the first of a series of experiments. The laboratory experimenter was blind to condition. The experimenter led participants to a separate cubicle and instructed them to complete the paper-and-pencil questionnaire.

The materials were the same as those in Study 1 except that the dependent variable was a hypothetical commons dilemma. In this commons dilemma, participants imagined a scenario in which they were the manager of a midsized industrial manufacturing plant. They read that all manufacturers reached an agreement to install filters to eliminate toxic gasses and to run these filters 60% of the time. Running a filter was costly for the manufacturing plant, but would be beneficial to society. To measure cooperative behavior, participants were asked to indicate what percentage of time they would operate the filters, indicated on an 11-point scale from 0 (labeled 0%) to 10 (labeled 100%).

After the main dependent variable, participants explained their decision and completed three secondary measures; they estimated (1) the percentage of other managers who would not cooperate, on the same 11-point scale; (2) the amount of environmental damage expected when the filters would be run less than the agreed 60% on an 11-point scale from 0 (none) to 10 (a great amount); and (3) the likelihood of getting caught when operating the filters less than 60% of the time on an 11-point scale from

0 (= *impossible*) to 10 (= *certain*). Finally, participants completed the seven self-presentation items from the Self-Monitoring scale (Lennox & Wolfe, 1984) and a set of demographic measures.

### Known Differences From Original Study

The only known difference compared to the original study is that we ran this study in a laboratory at a Dutch university instead of a USA university.

### Results

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analysis to test the effect writing about one's own (im)moral traits on cooperation (the amount of time participants were willing to run the filters). There were no significant differences between the conditions on cooperative behavior (Table 1).<sup>3</sup> Furthermore, there were no effects on the secondary variables (Table 2). The results of an additional regression model including gender, age, and ethnicity indicated that none of these demographic variables predicted cooperative behavior (all  $ps \geq .257$ ). Self-monitoring did not moderate the observed effects ( $p = .787$ ).

### Exploratory Analysis

We noticed that 48.5% of the participants violated the recall instructions and did not write about their own traits or used the words in a negating way. When we only used a post hoc selection of those who actually wrote about their own positive traits ( $N = 42$ ), there was still no difference on cooperative behavior between the positive trait stories about oneself and the neutral control condition ( $p = .197$ ).

## Study 3 – Replication of Sachdeva et al.'s (2009) Study 1 and Study 3 With a General US Population Sample on MTurk

### Participants

Whereas in Study 1 and Study 2, we based our sample size on a power analysis using the original studies, for Study 3 we did so based on the effect size of self-licensing that we obtained in the preliminary data of our meta-analysis ( $d = 0.26$ ) (Blanken et al., 2014). We calculated with G\*Power (Faul et al., 2009) that we would need at least 918 participants in our study to achieve 95% power to find a self-licensing effect. The sample was recruited on MTurk.

<sup>3</sup> A nonparametric independent-samples Kruskal-Wallis test (which controls for the not normally distributed data), also showed no effect,  $H(2) = 2.87, p = .238$ .

Table 1. Means, standard deviations, sample sizes, and test statistics for dependent variables in all studies

Dependent variable	Positive trait <i>M (SD)</i>	Neutral trait <i>M (SD)</i>	Negative trait <i>M (SD)</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Study 1, <i>N</i> = 105	35	36	34			
Donation amount <sup>a</sup>	2.89 (3.64)	2.78 (3.83)	2.35 (3.28)	0.21	.810	.004
Study 2, <i>N</i> = 150	49	49	52			
Cooperative behavior	6.29 (1.14)	5.88 (1.49)	5.84 (1.37)	1.70	.187	.023
Study 3, <i>N</i> = 940	306	326	308			
Donation amount	4.52 (2.91)	4.60 (3.11)	5.10 (3.18)	3.20	.041	.007
Cooperative behavior	6.27 (1.58)	6.39 (1.78)	6.21 (1.60)	0.78	.457	.002

Notes. For donation amount the answers could range from €0 to €10 (or participants could indicate another amount). For cooperative behavior, the answers indicate how long participants would choose to do the costly cooperative act in the scenario (run filters), on a range from 0 (*the least cooperative*) to 10 (*the most cooperative*). The statistical test of Study 3 is the main effect, controlling for order effects. Effect sizes and Confidence Intervals for the moral licensing and moral cleansing effects can be found in the Forest plots (Figures 1 and 2). <sup>a</sup>For the main dependent variable, participants were asked to indicate how much money they wanted to donate to a good cause. If they did not answer this question, we interpreted their response as €0. When these participants were excluded from the analysis, the results did not differ,  $F(2, 81) = 0.29$ ,  $p = .747$ ,  $\eta_p^2 = .007$ .

Table 2. Means, standard deviations, and test statistics for secondary measures in Studies 2 and 3

	Positive trait <i>M (SD)</i>	Neutral trait <i>M (SD)</i>	Negative trait <i>M (SD)</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Study 2, <i>N</i> = 150	49	49	52			
Expected cooperative behavior of others	4.64 (2.56)	4.57 (2.55)	4.28 (2.74)	0.26	.770	.004
Estimated likelihood of getting caught	5.92 (1.80)	5.61 (1.85)	5.62 (1.75)	0.47	.624	.006
Negative consequences for the environment	7.22 (1.62)	6.90 (1.95)	6.62 (1.88)	1.41	.248	.019
Study 3, <i>N</i> = 940	306	326	308			
Expected cooperative behavior of others	4.72 (2.91)	4.89 (3.08)	4.51 (2.86)	1.30	.273	.003
Estimated likelihood of getting caught	6.91 (2.07)	7.09 (2.14)	7.05 (2.02)	0.63	.535	.001
Negative consequences for the environment	5.84 (2.04)	5.98 (2.11)	5.93 (1.99)	0.33	.721	.001

Notes. For expected cooperative behavior of others, answers could range from 0 (the least cooperative) to 10 (the most cooperative). For estimated likelihood of getting caught, answers could range from 1 (impossible) to 10 (certain). For negative consequences for the environment, answers could range from 1 (none) to 10 (a great amount).

We included an instructional manipulation check to prevent inattentive participants from starting the study (see Oppenheimer, Meyvis, & Davidenko, 2009). Participants were asked to provide an answer to three neutral questions about stories and were explicitly instructed to answer “five” on the first question, and “seven” on the second and third question. Participants who did not follow these instructions ( $N = 160$ ) could not participate in our study. Our final sample consisted of 940 participants (449 males and 491 females,  $M_{\text{age}} = 33.41$ ) who participated in exchange for \$1.80.<sup>4</sup> All participants were randomly assigned to the positive trait condition ( $N = 306$ ), the negative trait condition ( $N = 308$ ), or the neutral control condition ( $N = 326$ ).

## Materials and Procedure

Participants completed the study materials via the Qualtrics survey program. Participants could subscribe to participate in our study entitled “writing style and several questions” if

they had an MTurk approval rate that was higher than 95% and if they lived in the US.

After finishing writing the stories with the positive traits, negative traits, or neutral words, participants answered the filler questions and both dependent measures from Sachdeva et al.’s (2009) Study 1 (donation amount) and Study 3 (cooperative behavior) in a counterbalanced order. Subsequently, participants completed the self-presentation items from the Self-Monitoring scale and a set of demographic measures.

## Known Differences From Original Study

The study was conducted online. We made two slight changes to these materials to increase the credibility of the online study. First, for the cover story, we instructed participants that the study was about general writing styles instead of handwriting, as the latter would not be believable in an online study. Second, we changed the donation

<sup>4</sup> We set the target higher than 918 to ensure a minimum of 918 valid participants after data exclusion.



measure. We told participants that 10 of them would be randomly selected to win an additional \$10 MTurk worker bonus. They were then asked that if they were one of the winners, would they be willing to donate a portion of this bonus to a cause of their choice from a list (cancer research, animal rights, ending world hunger, environmental preservation, human rights, veteran's affairs, or other). Participants selected a cause and indicated the amount they would donate ranging from \$0 to \$10 (or more).

## Results

### Donations

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analyses to test the effect of writing about (im)moral traits on how much participants would want to donate to a good cause. We controlled for order effects by including the order in which the two dependent variables were presented as a separate independent variable in the model. Order did not affect the donation amount,  $F(1, 934) = 0.78$ ,  $p = .378$ ,  $\eta_p^2 = .001$ , nor was there an interaction effect of order with the manipulation of what words participants wrote about,  $F(2, 934) = 0.42$ ,  $p = .656$ ,  $\eta_p^2 = .001$ .

As Table 1 shows, there was a main effect of moral identity condition on donation amount.<sup>5</sup> Post hoc Tukey tests indicated that participants in the negative trait condition donated more money than participants in the positive trait condition ( $p = .044$ ) and participants in the neutral control condition ( $p = .020$ ). There was no difference in donation amount between participants in the positive trait condition and participants in the neutral control condition ( $p = .729$ ). Thus, we did not find a moral licensing effect, but we did observe a moral cleansing effect – the recall of negative traits increased subsequent moral behavior. Self-monitoring did not moderate the observed effects.

Of the demographic variables gender, age, education level, family income, and ethnicity, only age significantly influenced donation amount ( $\beta = .11$ ,  $t(930) = 3.36$ ,  $p < .001$ ). When we included age as a covariate to the effect of the manipulation on donation amount, the effect of the manipulation remained significant,  $F(2, 932) = 3.15$ ,  $p = .043$ ,  $\eta_p^2 = .007$ .

### Cooperative Behavior

Next, we conducted Sachdeva et al.'s (2009) analyses to test the effect of moral identity condition on cooperation in a hypothetical commons dilemma. The order in which the dependent variables were presented did affect cooperative behavior,  $F(1, 934) = 11.20$ ,  $p = .001$ ,  $\eta_p^2 = .012$ , with participants who first completed the donation depen-

dent variable displaying slightly more cooperative behavior ( $M = 6.47$ ,  $SD = 1.77$ ) than participants who first completed this cooperative behavior dependent variable ( $M = 6.11$ ,  $SD = 1.52$ ). The interaction between moral identity and order was not significant,  $F(2, 934) = 0.83$ ,  $p = .438$ ,  $\eta_p^2 = .002$ . We do not know why this order effect exists, but for the current study it is mainly important that we control for this possible influence by adding it as a factor in the analyses. As Table 1 shows, there was no main effect of moral identity on cooperative behavior,<sup>6</sup> nor on the secondary variables (see Table 2). Again, self-monitoring did not moderate the observed effects.

Of the demographic variables, only one of the ethnicity dummy variables significantly influenced cooperation (with African Americans cooperating less than others,  $\beta = -.19$ ,  $t(930) = -3.11$ ,  $p = .002$ ). When including ethnicity as a covariate, there was still no effect of moral identity condition on cooperative behavior,  $F(2, 933) = 0.81$ ,  $p = .447$ ,  $\eta_p^2 = .002$ .

### Exploratory Analyses

We noticed that 43.6% of the participants violated the recall instructions and did not write about their own traits or used the words in a negating way. Using solely the coded stories about oneself in our analyses, there was a main effect of moral identity condition on donation amount ( $p = .020$ ) with participants in the negative trait condition donating more money than participants in the positive trait condition ( $p = .017$ ) and in the neutral control condition ( $p = .009$ ). There was no main effect of moral identity condition on cooperative behavior ( $p = .495$ ).

## General Discussion

We made three attempts to replicate the findings of Sachdeva et al. (2009) on moral licensing, with samples based on pre-calculated power and preplanned analyses. In the first two replication attempts using student samples, the data did not confirm the original results. In our third replication attempt using a general population sample the data did not confirm the moral licensing effect. We did, however, find support for the moral cleansing effect on one of the two dependent variables in Study 3, but not in Studies 1 and 2.

### Current Status of the Moral Licensing Effect

We conducted a meta-analysis of this moral licensing effect by including both the original Studies 1 and 3 by Sachdeva et al. and the three current replication attempts, using the metafor package of Viechtbauer (2010). For our Study 3,

<sup>5</sup> A nonparametric independent-samples Kruskal-Wallis test (which controls for the skewness of the data), found a similar effect,  $H(2) = 5.85$ ,  $p = .054$ .

<sup>6</sup> A nonparametric independent-samples Kruskal-Wallis test (which controls for the not normally distributed data), also found a nonsignificant effect,  $H(2) = 2.68$ ,  $p = .713$ .

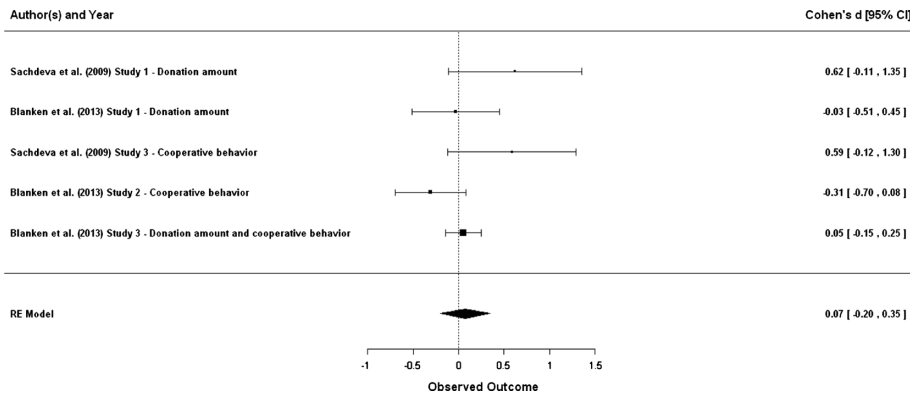


Figure 1. Forest plot including all comparisons between the moral licensing and neutral control conditions of the original studies by Sachdeva et al. (2009) and our replication attempts.

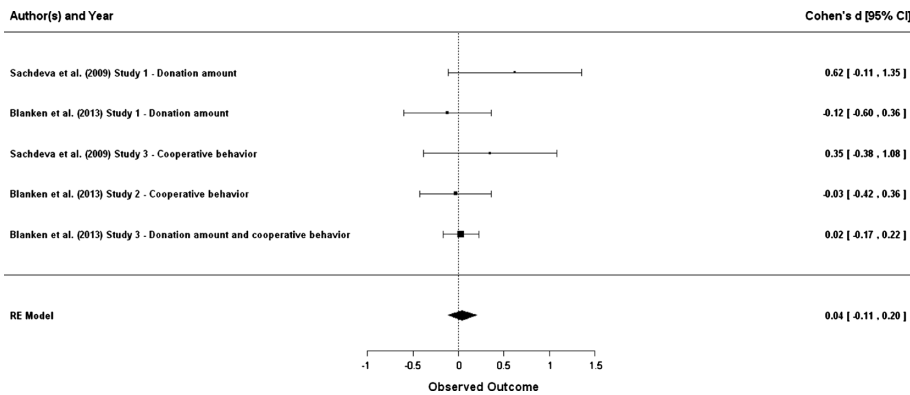


Figure 2. Forest plot including all comparisons between the moral cleansing and neutral control conditions of the original studies by Sachdeva et al. (2009) and our replication attempts.

we used the average effect size of the two dependent variables. The random effects meta-analysis including all five studies produced a mean effect size of moral licensing of  $d = 0.07$  ([ $CL_{95}$ ]  $-0.20$  to  $0.35$ ). There was thus no significant moral licensing effect across studies ( $z = 0.52, p = .603$ ). Figure 1 contains an overview of all moral licensing effect sizes (when compared to the neutral control conditions).

### Current Status of the Moral Cleansing Effect

We conducted a meta-analysis of this moral cleansing effect by including both the original Studies 1 and 3 by Sachdeva et al. and the three current replication attempts. The random effects meta-analysis including all five studies produced a mean effect size of moral cleansing of  $d = 0.04$  ([ $CL_{95}$ ]  $-0.11$  to  $0.20$ ). There was thus no significant moral cleansing effect across studies ( $z = 0.53, p = .593$ ). Figure 2 contains an overview of all moral cleansing effect sizes (when compared to the neutral control conditions). However, note that only a small number of participants in the moral cleansing condition of the replication studies actually wrote about themselves.

### Possible Limitations of Our Replication Attempts

Although we did our best to design direct replications of the original studies, differences are inevitable, and some of

those may be consequential for moderating the results. First, our Studies 1 and 2 used Dutch students not US students. There is no theoretical reason to expect different licensing effects for Dutch compared to US citizens, but our pilot test (see supplements) suggested that words in the positive moral trait condition were seen to be slightly more positive in the US than in the Netherlands. Even so, the words were evaluated very positively in both national samples. Study 3 used a US based sample, but this study differed on two aspects compared to the original study. It was conducted online instead of in the laboratory, and the manipulation involved donating a part of potential winnings instead of money out-of-pocket. We cannot rule out that these procedural differences were consequential, but there presently exists no theoretical reason or identification of these as boundary conditions on moral licensing.

### Conclusion

Although Sachdeva et al. (2009) theorized that moral licensing and moral cleansing should be considered jointly as being part of a moral self-regulation process, our three high-powered studies did not replicate the key moral licensing effect. Further, the meta-analytic result suggests that the present state of evidence with this paradigm is not different from a null effect. Sachdeva et al. (2009, p. 524) suggested that their findings showed that “moral-licensing and moral-

cleansing effects can act convergently as part of a moral self-regulation process.” Based on the present findings, we do not argue that the theory is incorrect, only that it lacks sufficient empirical support when using the Sachdeva et al. (2009) paradigm.

We suggest three concrete steps to clarify the effects of moral licensing on social judgment. First, the method used by Sachdeva et al. (2009) seems unlikely to elicit moral licensing, especially since many participants violated the recall instructions and did not write about their own traits or used the words in a negating way. This is a procedure-specific issue; it does not invalidate moral licensing more generally. Second, the meta-analysis of all licensing research suggests that the effect is relatively small (Blanken et al., 2014). Therefore, small sample studies are highly inadvisable as they would need to leverage chance to detect a result using null hypothesis significance testing. Third, because moral licensing and moral cleansing are theoretically distinct, it is important to use a neutral control condition to clarify the role of each in social judgment.

## Acknowledgments

This work was supported by a grant from the Center of Open Science. The authors gratefully thank Sonya Sachdeva and her co-authors for sharing all materials and information about the methodology and for her helpful comments. The authors declare no conflict-of-interest with the content of this article. Designed research: I.B., N.V., M.Z., M.H.C.M.; Performed research: I.B.; Analyzed data: I.B.; Wrote paper: I.B., N.V., M.Z., M.H.C.M. All materials, data, and the preregistered design are available at: <https://osf.io/3cmz4/>.



## References

- Aronson, J., Cohen, G. L., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 127–148). Washington, DC: American Psychological Association.
- Blanken, I., Van de Ven, N., & Zeelenberg, M. (2014). *A meta-analytic review of self-licensing*. Manuscript under review.
- Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, *6*, 907–919.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology*, *45*, 590–593.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University.
- Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology*, *201*, 107–112.
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, *37*, 701–713.
- Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, *43*, 357–365.
- Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology*, *46*, 1349–1364. doi: 10.1037/0022-3514.46.6.1349
- Mazar, N., & Zhong, C. B. (2010). Do green products make us better people? *Psychological Science*, *21*, 494–498.
- 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*, 344–357.
- Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. *Advances in Experimental Social Psychology*, *43*, 115–155.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43.
- Mukhopadhyay, A., Sengupta, J., & Ramanathan, S. (2008). Recalling past temptations: An information-processing perspective on the dynamics of self-control. *Journal of Consumer Research*, *35*, 586–599.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*, 523–528.
- Schlenker, B. R. (1980). *Impression management: The self-concept, social identity and interpersonal relations*. Monterey, CA: Brooks/Cole.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). New York, NY: Academic Press.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Received February 28, 2013

Accepted December 3, 2013

Published online May 19, 2014

Irene Blanken

Department of Social Psychology  
Tilburg University  
PO Box 90153  
5000 LE Tilburg  
The Netherlands  
E-mail [i.blanken@tilburguniversity.edu](mailto:i.blanken@tilburguniversity.edu)