Does Recalling Moral Behavior Change the Perception of Brightness?

A Replication and Meta-Analysis of Banerjee, Chatterjee, and Sinha (2012)

Mark J. Brandt, Hans IJzerman, and Irene Blanken

Tilburg University, The Netherlands

Abstract. Banerjee, Chatterjee, and Sinha (2012) recently reported that recalling unethical behavior led participants to see the room as darker and to desire more light-emitting products (e.g., a flashlight) compared to recalling ethical behavior. We replicated the methods of these two original studies with four high-powered replication studies (two online and two in the laboratory). Our results did not differ significantly from zero, 9 out of 10 of the effects were significantly smaller than the originally reported effects, and the effects were not consistently moderated by individual difference measures of potential discrepancies between the original and the replication samples. A meta-analysis that includes both the original and replication effects of moral recall on perceptions of brightness find a small, marginally significant effect (d = 0.14 $CL_{95} - 0.002$ to 0.28). A meta-analysis that includes both the original and replication effects of moral recall on preferences for light-emitting products finds a small effect that did not differ from zero (d = 0.13 $CL_{95} - 0.04$ to 0.29).

Keywords: morality, light, grounded cognition, embodiment

One recent addition to the literature on grounded cognition, specifically on conceptual metaphors of morality, examined how reminders of people's own morality or immorality can shape perceptions of lightness and darkness (Banerjee, Chatterjee, & Sinha, 2012; from here on referred to as "BCS"). BCS requested participants to recall a time they engaged in ethical or unethical behavior and then asked how light or dark the room was. They found that participants who recalled three unethical deeds and subsequently wrote about the most unethical deed perceived the room as darker (Study 1, N = 40) and being lit with fewer Watts (Study 2, N = 74) than participants who recalled three ethical deeds and subsequently wrote about the most ethical deed. These effects provide support for the idea that the abstract target domain of morality influences perceptions in the concrete source domain of light (see Figures 1 and 2 for effect sizes and confidence intervals; see Firestone & Scholl, 2014, for a recent alternative explanation for this effect). An intriguing addition to BCS Study 2 was the finding that people in the unethical condition compared to the ethical condition were more likely to prefer products that convey light (e.g., lamps, flashlights), presumably because they perceive their environment to be darker. That is, these studies have found that abstract thought (morality) shapes concrete experiences (perception of light) (the abstract \rightarrow

concrete causal direction). We aimed to replicate these studies as closely as possible.

The work by BCS builds on other studies that have linked immorality/morality with darkness/lightness (Frank & Gilovich, 1988; Sherman & Clore, 2009; Webster, Urland, & Correll, 2012; Zhong, Bohns, & Gino, 2010). These studies all suggest that the concrete source domain of color and light perception influences the abstract target domain of morality (the concrete \rightarrow abstract causal direction). These results are consistent with the linguists Lakoff and Johnson's (1999) suggesting that conceptual metaphors are unidirectional. That is, the learning of an abstract concept may co-occur with the concrete experience, but the concrete experience is not necessarily associated with the abstract concept. Concretely, this means that lightness/darkness should lead to an alteration in perceptions of morality, but priming of moral or immoral deeds should not lead to perceptions of lightness/darkness (for skepticism of this argument see IJzerman & Koole, 2011; IJzerman & Semin, 2010; for skepticism of this skepticism, see Lee & Schwartz, 2012; Slepian & Ambady, 2014). The studies by BCS are important for understanding the theoretical link of light and morality because the studies by BCS suggest that the morality-light association goes beyond such conceptual metaphors (cf. Lakoff & Johnson, 1999).

They indicate that the abstract concept of morality guides our processing of color/light information and influences our perception of light in our environment, thereby potentially suggesting Conceptual Metaphor Theory is incomplete, or that moral concepts are grounded in basic perceptual simulators (Barsalou, 1999; IJzerman & Koole, 2010; see also Hamlin, Wynn, & Bloom, 2007, for reasoning that may suggest this argument).

Methods

We replicated BCS Studies 1 and 2 using the original methods from BCS that were provided to us by the original authors. The details of our methods, the precise differences between our replications and the original studies, and our sample size justifications and planning can be found in online supplemental material and in the original preregistration of the studies. Because the methods of the two original studies are largely the same, with the exception of the dependent variables, we simultaneously describe our four replication studies and note where they deviate from one another.

Procedure and Measures

All of the participants completed the studies on computers. Participants in our online samples completed the study on their own computer. The laboratory samples completed the study on computers in individual cubicles (see a photo of a cubicle and a video simulation from one week of our replication of Study 2 in the supplemental materials). All of the measures from the original study were included in the replications. In the replication of Study 1, participants described in detail an ethical or an unethical deed from their past, completed filler items about the room they were in, and made judgments of the brightness in the room on a 7-point scale (1 = not bright at all, 7 = very bright). In the replication of Study 2, the procedure was the same, except following the filler items participants rated their preferences (1 = not at)all desirable, 7 = very desirable) for light-emitting (lamp, candle, flashlight) and filler (jug, crackers, apple) products before estimating the brightness of the room in watts. The brightness judgments (Studies 1 & 2) and preference for light-emitting products (Study 2) were the primary dependent variables. We also included several additional measures of demographic information, religiosity, political ideology, and moral-self identification at the very end of the study to test possible moderators that may explain differences between the original and replication samples, and to maximize chances to obtain an effect at all.

Participants

For each of the original two studies reported by BCS we conducted two replication attempts, one online via MTurk where participants received \$0.50 and one in our laboratory at Tilburg University in the Netherlands where participants

received course credit or 5 Euros (see below). The final sample sizes and basic demographic information from both the original study and our replication studies are in Table 1. The sample sizes reported in Table 1 are the largest sample sizes available for the study; however, given that some participants did not complete all of the measures the precise degrees of freedom vary depending on the analysis.

In the online studies we aimed for Ns of 496 and 510 for Studies 1 and 2, respectively. In the laboratory studies we aimed for Ns of 126 and 130 for Studies 1 and 2, respectively. We aimed for these sample sizes because they would give us 95% power given the effect sizes reported by BCS (and assuming that the effect sizes in the online studies would be 50% weaker than in the original studies that were conducted in the laboratory). Although we followed the data collection protocol and stoppage rules outlined in our preregistration for laboratory Study 1 and online Studies 1 and 2, we fell short of our sample size goals because there were more participants than expected in our online samples who did not follow directions or completed the study outdoors. In our laboratory study, we did not collect the expected sample size because we were unfortunate to collect data during a "slow" laboratory week. For laboratory Study 2, we collected data for 1 week (as specified in our preregistration) and participants were compensated with partial course credit; however, we did not have nearly a sufficient number of participants (N = 66). Therefore we collected data for 2 additional weeks and participants were compensated with $\in 5$ (N = 55). Analyses that take the "week of data collection" into account do not alter the conclusions we report below. Each of our studies still had a high amount of power to detect effects of the size reported by BCS. The achieved power was above typically recommended power levels (e.g., .80 by Cohen, 1988) (Lowest Achieved Power Online Study 1 = .94, Online Study 2 = .93, Laboratory Study 1 = .90, Laboratory Study 2 = .90).

Results

Our confirmatory analyses replicated the analyses reported in BCS (i.e., independent-sample *t*-tests comparing experimental conditions) and can be found in Table 2. In the online studies, the effects of the experimental conditions on all of the primary dependent variables were nonsignificant (all *t*'s < |1.28|, all *p*'s > .20). Similarly, in the laboratory studies the effects of the experimental conditions on all of the primary dependent variables were nonsignificant (all *t*'s < |0.59|, all *p*'s > .55). With the exception of the estimation of brightness in the online version of Study 1, all of the effect sizes were significantly smaller in the replication studies than the original study (see Table 2).

Exploratory Analyses

We tested to see if age, gender, ethnicity (online studies), education (online studies), income (online studies), importance

247

	BCS		B&I: Online		B&I: Laboratory	
	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
Ν	40	74	475	482	100	121
M age (SD)	NR	NR	28.8 (9.5)	29.2 (9.5)	19.6 (2.2)	20.0 (2.3)
Gender (M/F)	NR	NR	173/301 ^a	169/313 ^a	27/73	44/77
Population	"participants at a large public university"	NR. Assumed to be the same as Study 1	MTurk workers	MTurk workers	Dutch university students	Dutch university students
Population location Study setting	United States Computer in laboratory	United States Computer in laboratory	United States Online	United States Online	Netherlands Computer in individual laboratory cubicle	Netherlands Computer in individual laboratory cubicle

Table 1. Final sample sizes and demographic information in the original and replication studies

Notes. NR = not reported. ^a1 person did not report their gender.

Table 2. Means, standard deviations, and effect sizes for the original and replication studies

	BCS		Replication: Online		Replication: Laboratory		
	Unethical prime	Ethical prime	Unethical prime	Ethical prime	Unethical prime	Ethical prime	
Study 1							
Perceived brightness M	4.71	5.3	4.34	4.51	4.79	4.66	
Perceived brightness SD	0.85	0.97	1.52	1.47	1.09	1.19	
Perceived brightness d		55 _a .1		.12 _{ab}		11 ^b	
Achieved power	.52		>.99		.90		
Study 2							
Estimated watts M	74.3	87.6	296.12	443.88	130.44	135.91	
Estimated watts SD	26.85	7.40	1,095.43	3,991.04	152.03	192.64	
Estimated watts d	.64 _a		.05 _b		.03 _b		
Achieved power	.7	78	>.9	99		0.9	
Lamp preference M	4.16	2.34	4.02	4.07	3.44	3.62	
Lamp preference SD	1.70	1.15	1.59	1.64	1.74	1.47	
Lamp preference d	1.2	23 _a	()3 _b		11 _b	
Achieved Power	>.	99	>.9	99		>.99	
Candle preference M	3.62	2.37	3.34	3.33	4.30	4.32	
Candle preference SD	1.83	1.16	1.76	1.80	1.67	1.51	
Candle preference d	eference d .79 _a		.003 _b		01_{b}		
Achieved power	.9	02	>.9	99		.99	
Flashlight preference M	4.33	2.35	3.26	3.43	2.67	2.80	
Flashlight preference SD	1.71	1.15	1.75	1.78	1.58	1.39	
Flashlight preference d	1.3	33 _a	1	10 _b		09 _b	
Achieved power	>.	99	>.9	99		>.99	

Notes. Achieved power for the replications is the achieved power based on the effect sizes reported in BCS (without the adjustments made for online vs. laboratory studies, see main text and preregistration). Effect sizes within the same row with different subscripts are significantly different from one another p < .05. BCS effect sizes are the effect sizes reported in the paper. Effect sizes from our studies were calculated with Becker's effect size calculator: http://www.uccs.edu/~lbecker/. Differences between the effect sizes were computed by first computing the equivalent *r*-value for each *d*-value and then computing a *z*-score for differences in correlation coefficients (http://vassarstats.net/rdiff.html).

of morality to the self, religiosity growing up, current religiosity, and political ideology moderated the effect of the experimental manipulations on any of the primary dependent measures. There were 3 of 80 moderation effects were possible. None of the significant differences were observed consistently across studies.



Figure 1. Cohen's *d*, 95% confidence intervals, and estimate of overall effect size from a random effects meta-analytic model for the perceptions of brightness. PB = perceived brightness, EW = estimated watts.

Meta-Analyses

The results of any one study, including high-powered replication studies, could be the result of chance. Similarly, the original studies may have uncovered robust effects, but by chance estimated the effect sizes as much larger than the true effects. Therefore, to gain a more precise understanding of the effects we conducted two meta-analyses (one on the brightness judgments and one on the desirability of light-emitting products) including the original studies, the replication attempts reported here, our own previous replication attempts (Brandt, IJzerman, & Blanken, 2013), and two other recent published replication attempts of Study 1 of BCS (Firestone & Scholl, 2014). With the information we collected, we were also able to test whether the effect was more robust online or in the laboratory and whether it was more likely in the United States or in the Netherlands. Although not specified in our preregistration, we also tested whether the research laboratory where the study was conducted affected the obtained effect sizes. All analyses were conducted using the metaphor package for the R program (Viechtbauer, 2010).

Meta-Analysis on the Effects of Experimental Condition on Brightness Judgments

We first conducted a meta-analysis to derive the overall mean effect size of experimental condition on brightness judgments (N = 11). The random effects meta-analysis produced a mean effect size of d = 0.14 ([CL₉₅] -0.002 to 0.28). There was a marginal effect of experimental condition across all the studies on brightness judgments (z = 1.93, p = .054). Figure 1 provides a forest plot of the effect sizes of the brightness judgments across studies. The effect of experimental condition on brightness judgments did not differ for participants from the US (M effect size = 0.17, SE = 0.09) versus participants from the Netherlands

(*M* effect size = 0.08, SE = 0.16), QM(2) = 3.63, p = .16. The effect was larger for studies conducted in the laboratory (*M* effect size = 0.24, SE = 0.12, p = .05) than for studies conducted online (*M* effect size = 0.08, SE = 0.12, p = .36), QM(2) = 4.85, p = .04.

Our exploratory analysis on research laboratory where the study was conducted yielded an overall significant effect, QM(4) = 18.45, p = .001, with studies conducted in the Banerjee Laboratory (*M* effect size = 0.64, SE = 0.20) and in the Firestone Laboratory (*M* effect size = 0.42, SE = 0.16) showing significant effects in the positive direction (p = .001 and p = .1, respectively) and studies conducted in the cubicles by the Brandt Laboratory (*M* effect size = -0.04, SE = 0.14) and online by the Brandt Laboratory (*M* effect size = 0.04, SE = 0.05) showing no significant effects (p = .8 and p = .3, respectively).

Meta-Analysis on the Effects of Experimental Condition on the Desirability of Light-Emitting Products

Next, we conducted a meta-analysis to estimate the overall mean effect size of experimental condition on the desirability of light-emitting products (N = 15). The random effects meta-analysis produced a mean effect size of d = 0.13 $([CL_{95}] - 0.04$ to 0.29). There was no significant effect of experimental condition across all studies on brightness judgments (z = 1.53, p = .13). Figure 2 provides a forest plot of the effect sizes of the desirability of light-emitting products across studies. The effect of experimental condition on the desirability of light-emitting products did not differ for participants from the US (M effect size = 0.25, SE = 0.15) versus participants from the Netherlands (M effect size = 0.01, SE = 0.18, QM(2) = 2.81, p = .25. The effect was larger for studies conducted in the laboratory (M effect size = 0.33, SE = 0.13, p = .01) than for studies conducted online (M effect size = -0.09, SE = 0.15, p = .56, QM(2) = 6.37, p = .04.



Figure 2. Cohen's *d*, 95% confidence intervals, and estimate of overall effect size from a random effects meta-analytic model for the preferences for light-emitting products. LP = lamp preference, CP = candle preference, FP = flash-light preference.

Our exploratory analysis on research laboratory where the study was conducted yielded an overall significant effect, QM(3) = 56.22, p < .001, with studies conducted in the Banerjee laboratory (M effect size = 1.10, SE = 0.15) showing significant effects in the positive direction (p < .001) and studies conducted in the Brandt laboratory (M effect size = -0.03, SE = 0.06) and online (Meffect size = -0.06, SE = 0.05) showing no significant effects (p = .64 and p = .25, respectively).

Discussion

Despite conducting high-powered replication studies of BCS, we were unable to replicate the original effects in our own replication studies. Recalling ethical or unethical behavior did not have an effect on the estimated brightness of the room, the estimated watts of light in the room, or the preference for light-emitting products. A meta-analysis of available effect sizes of moral recall on perceptions of brightness indicated that on average there is a marginally significant effect that tends to be larger when tested in a laboratory setting. The meta-analysis on preferences for light-emitting products did not reveal any effect of moral recall on product preferences, suggesting that this effect may be less robust. This effect was also moderated by whether the study was in the laboratory or online, with a significant effect on average when conducted in the laboratory. Overall, we believe that there is still much to be learned about the robustness of the effect of moral recall on the perception of light. The replications and meta-analysis reported here suggest that the effect is not robust; however, two independent laboratories have observed the effect.

At this stage we think it is important to try and understand why BCS and (Firestone & Scholl, 2014) were able to detect the effect and we were not. It may be that subtle aspects of the procedure, whether in the formatting of the study, the wording of the consent form, or some other feature is essential for the effect and was different between the available studies. This is a clear possibility because Firestone and Scholl (2014) found the effect in the same online population (i.e., MTurk) where we collected our online data, even though a moderator analysis suggested that online studies produced weaker effects on average. Similarly, it seems unlikely that our Dutch laboratory studies are a cause for concern because others have detected links between immorality/morality and darkness/lightness in Dutch samples (Lakens, Semin, & Foroni, 2012), classic social psychological effects have been replicated in our Tilburg laboratories (Klein et al., 2014), and we also detected a similar null effect with online American samples. This led us to consider the "laboratory group" that conducted the study as a potential moderator in the meta-analyses. These moderation analyses suggest that something about the particular laboratory that conducted the study may be driving the effect. This could be something about the precise display of the stimuli within the experimental program or other aspects of the experimental setting and presentation.

One specific direction for future work is to explore the differences between our online replication attempts and the two attempts reported by Firestone and Scholl (2014). We both collected data from the MTurk population; however, subsequently we have learned that whereas we used an 80% approval rating for MTurk workers (an indicator of worker quality), Firestone and Scholl used a more stringent 95% approval rating for MTurk workers. The type of samples drawn from the MTurk population may significantly differ between these two approval rate levels and this

may explain the differences between our online studies and the Firestone and Scholl online studies. It should be noted, however, that this does not explain the discrepancy between our laboratory replications and the laboratory replications of BCS.

A second direction researchers could explore is the fact that both in our laboratory studies and our online replication studies participants estimated a large range of watts as lighting the room. For example, the standard deviations for our laboratory study, where all of the participants were in individual cubicles illuminated by the same 66 Watt fluorescent light, were larger than 150 Watts. BCS, on the other hand, estimated the standard deviation to be about one sixth of the size. This may indicate more knowledge or attention to the light in the room by BCS's participants compared to our participants in the laboratory studies. In light of the issues and potential causes for our null results discussed above, future investigations into the nature of the effect of moral recall on perceptions of brightness should keep careful records of the differences between the original and replication study on more basic issues in regard to stimulus presentation, experimental context, and attention to one's surroundings to potentially find the key to the effect (cf. Brandt et al., 2014; Cesario, 2014).

In conclusion, we are hesitant to proclaim the effect a false positive based on our null findings, or a true success based on the marginally significant meta-analytic effect. Instead we think that scholars interested in how morality is grounded should be hesitant to incorporate the studies reported by BCS into their theories until the effect is further replicated and a possible explanation for the discrepancy between our findings and the original findings is identified *and* tested. Until answers to these questions are available it appears that the possibility of an abstract concept (morality) changing people's perception of something more concrete (perception of light) will remain just that, a possibility.

Acknowledgments

We thank Ellen Evers for helping collect data for the Study 2 replication in the laboratory. We especially thank Banerjee, Chatterjee, and Sinha for providing their original materials and to both the Banerjee group and Firestone and Scholl for providing additional study details for the metaanalysis. We report all data exclusions, manipulations, and measures, and how we determined our sample sizes. The authors declare no conflict-of-interest with the content of this article. Designed research: M. B., H. I., Performed research: M. B. Analyzed data: M. B. Conducted and Reported Meta-analysis: I. B. Wrote paper: M. B., H. I. The replications reported in this article were supported by a grant from the Center for Open Science awarded to the three authors (Grant # 2013-002). All materials, data, syntax, supplemental information, supplemental analyses, and the preregistered design are available here: http:// osf.io/seqgd/.



References

- *Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science*, 23, 407–409.
- Barsalou, L. W. (1999). Perceptual symbols system. *Behavioral* and Brain Sciences, 22, 577–660.
- *Brandt, M. J., IJzerman, I., & Blanken, I. (2013). *Replication of Banerjee, Chatterjee, and Sinha (2012)* Unpublished replication report and data. https://osf.io/jpfsi/.
- 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.
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9, 40–48.
 *Firestone, C., & Scholl, B. J. (2014). "Top-down" effects
- *Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, 25, 38–46.
- Frank, M. G., & Gilovich, T. (1988). The dark side of self-and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, 54, 74–85.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557–559.
- IJzerman, H., & Koole, S. L. (2010). From perceptual rags to metaphoric riches – Bodily, social, and cultural constraints on sociocognitive metaphors: Comment on Landau, Meier, and Keefer (2010). *Psychological Bulletin*, 137, 355–361.
- IJzerman, H., & Semin, G. R. (2010). Temperature perceptions as a ground for social proximity. *Journal of Experimental Social Psychology*, 46, 867–873.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B. Jr., Bahník, Š., Bernstein, M. J., ... Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, 45, 142–152.
- Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, 141, 584–594.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York, NY: Basic Books.
- Lee, S. W., & Schwarz, N. (2012). Bidirectionality, mediation, and moderation of metaphorical effects: The embodiment of social suspicion and fishy smells. *Journal of Personality and Social Psychology*, 103, 737–749.
- Sherman, G. D., & Clore, G. L. (2009). The color of sin: White and black are perceptual symbols of moral purity and pollution. *Psychological Science*, 20, 1019–1025.
- Slepian, M. L., & Ambady, N. (2014). Simulating sensorimotor metaphors: Novel metaphors influence embodied cognition. *Cognition*, 130, 309–314.

Indicates studies included in the meta-analysis.

- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Webster, G. D., Urland, G. R., & Correll, J. (2012). Can uniform color color aggression? Quasi-experimental evidence from professional ice hockey. *Social Psychological* and Personality Science, 3, 274–281.
- Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and selfinterested behavior. *Psychological Science*, *21*, 311–314.

Received February 28, 2013 Accepted December 17, 2013 Published online May 19, 2014

Mark J. Brandt

Department of Social Psychology Room P06 P.O. Box 90153 Tilburg University 5000 Tilburg The Netherlands E-mail m.j.brandt@tilburguniversity.edu