IN PRESS - *SOCIAL PSYCHOLOGY*

The Relationship Between Immorality and Cleansing: A Meta-Analysis of the Macbeth Effect

Jedidiah Siev,[1*] Shelby E. Zuckerman,[2] and Joseph J. Siev[3]

[1]Swarthmore College

[2]VA Boston Healthcare System

[3]The Ohio State University

* Correspondence may be directed to:

Jedidiah Siev, Ph.D.
Department of Psychology
Swarthmore College
500 College Ave
Swarthmore, PA 19081
jsiev1@swarthmore.edu
1-610-328-8418

Abstract

In a widely-publicized set of studies, participants who were primed to consider unethical events preferred cleansing products more than did those primed with ethical events (Zhong & Liljenquist, 2006). This tendency to respond to moral threat with physical cleaning is known as the Macbeth Effect. Several subsequent efforts, however, did not replicate this relationship. The present manuscript reports the results of a meta-analysis of 15 studies testing this relationship. The weighted mean effect size was small across all studies (g = 0.17, 95% CI [0.04 – 0.31]), and non-significant across studies conducted in independent laboratories (g = 0.07, 95% CI [-0.04 – 0.19]). We conclude that there is little evidence for an overall Macbeth Effect; however, there may be a Macbeth Effect under certain conditions.

Keywords: Macbeth Effect, morality, cleansing, metaphor, embodiment

The Relationship Between Immorality and Cleansing: A Meta-Analysis of the Macbeth Effect

In a widely publicized and influential set of studies, Zhong and Liljenquist (2006) investigated the psychological link between morality and physical cleansing. They posited a bidirectional relationship in which moral threat motivates physical cleansing, and cleansing alleviates moral threat whether or not the motivation is to do so. The former of these effects has been dubbed the Macbeth Effect, a reference to the Shakespearean Lady Macbeth's attempt to soothe her conscience by washing her hands after she murdered King Duncan. The authors found that participants who were primed to consider unethical events generated more words related to cleansing, rated cleansing products as more desirable, and chose cleansing-related products more often than did participants primed with ethical events. This set of studies has been impactful, having been cited more than 800 times according to Google Scholar.

Several efforts at replication, however, did not find a relationship between immorality and cleansing, even using similar methods and procedures, larger samples, and diverse participants. Historically, there has been a longstanding publication bias against non-significant findings. As a result of this bias, in combination with the likelihood that initial studies have small samples, initial published studies are likely to have particularly large effect sizes (e.g., Ioannidis, 2008; Levine, Asada, & Carpenter, 2009). Other factors, including inadequate specification in an initial published report of the conditions necessary to obtain an effect, may also contribute to declining effect sizes in subsequent studies (Protzko & Schooler, 2017).

More generally, there is a widely-acknowledged replication crisis in the field of psychology, and studies of unconscious priming effects are no exception (e.g., Doyen, Klein, Simons, & Cleeremans, 2014). This crisis has been influenced by systemic incentives that discourage replication efforts and decrease the likelihood that non-significant findings are

published (for reviews of the history of this crisis and factors that influence it, see e.g., Earp, 2017; Earp & Trafimow, 2015). These incentives, such as preference for positive results and novelty (e.g., Nosek, Spies, & Motyl, 2012), have been changing with growing awareness of their consequences. Across much of the field, both null findings and replication efforts are increasingly valued and more frequently published in recent years. Therefore, it is of particular interest to quantify the magnitude of effects that are influential and appeared large in initial publications, but have failed so far to replicate.

In the current investigation, we seek to quantify the magnitude of the Macbeth Effect. We report the results of a random-effects meta-analysis of the existing data, including two as-yet unpublished studies of our own and one unpublished study by the first author of the original paper. We also draw upon relevant theoretical literatures in an attempt to encourage a more granular description of the psychological mechanisms implicated in a Macbeth Effect, to propose boundary conditions suggested by an integrated view of the data, and to point toward promising avenues for future empirical study.

Method

Study Selection

Studies examining the Macbeth Effect were located via *PsycINFO* with search terms "Macbeth Effect," "Macbeth," and "morality and cleansing"; by examining journals and references listed in related publications; by searching PsychFileDrawer.org; and by soliciting unpublished data via the listserv of the Society for Personality and Social Psychology, as well as directly contacting the authors of the studies included in the meta-analysis. Two unpublished replication attempts conducted by the first author were also included. Studies were included if they examined the effects of an immoral versus moral prime on cleansing-related preferences or

behaviors. Studies that examined the effects of cleansing on morality were excluded because that question and whether primes related to moral threat create a preference for cleansing are conceptually and empirically distinct.

In order to achieve the most parsimonious test of the Macbeth Effect, we included only direct or slightly modified replications of Zhong and Liljenquist's (2006) first three studies. In Study 1, participants recalled an unethical or ethical deed and then completed a hangman-style word completion task in which there were cleansing-related and -unrelated solutions (e.g., W--H). The dependent variable was the number of cleansing-related solutions. In Study 2, participants copied an unethical or ethical story and then rated their preferences for a number of products, including cleansing products. The dependent variable was the desirability of cleansing products. In Study 3, participants completed the same recall task as in Study 1 and then were offered a gift of either a pencil or antiseptic wipe. The dependent variable was the likelihood of choosing the antiseptic wipe.

Studies were deemed direct replication efforts if (a) they used a between-subjects experimental design, (b) participants in both conditions received a prime that was used in one of the studies by Zhong and Liljenquist (2006) or a close adaptation thereof, (c) and the dependent measure related to cleansing preference was implicit or not recognized by the participants. Studies were excluded if participants received an additional task manipulation between the prime and dependent measure. Small methodological variations were allowed in the absence of specific evidence that they may moderate outcome. Therefore, studies were included with international participants, different languages, and both computerized and pen-and-paper tasks. The decision to focus on direct replication efforts is consistent with Simons' (2014) observation that direct replication by other laboratories is "the best (and possibly the only) believable evidence for the

reliability of an effect" (p. 76; on the necessity of considering studies conducted outside the original laboratory, see also Brandt et al., 2014). Others have noted that conceptual replication efforts are of limited value without direct replication because failed efforts will be attributed to methodological variation and therefore will not challenge beliefs about the original effect (Doyen et al., 2014).

See Table 1 for a list of included and excluded studies, as well as methodological characteristics vis-à-vis the original studies.

Data Analysis

The weighted mean effect size was computed using a random-effects model, which included standardized mean difference effect sizes from 15 studies conducted with 1746 participants from three different continents, reported in 6 publications and 3 unpublished datasets. Supplemental data or materials were provided by the authors for all three studies by Earp et al. (J. A. C. Everett, personal communication, June 12, 2014), Zhong and Liljenquist's first study (C.-B. Zhong, personal communication, May 16, 2014), Lee and Schwarz (S. W. S. Lee, personal communication, May 20, 2014), Zhong's unpublished study (C.-B. Zhong, personal communication, November 7, 2017), and D'Olimpio and Mancini (F. D'Olimpio, personal communication, May 19, 2016). Analyses were conducted using Comprehensive Meta-Analysis software, which uses the DerSimonian and Laird (1986) method of weighting. The dataset is available on the OSF website (https://osf.io/twsxf).

Standardized mean difference. Standardized mean difference effect sizes (Hedges' *g*) were calculated such that positive numbers denote greater cleansing preference in the unethical than ethical condition. For the word completion results and product ratings, effect sizes were computed from means and standard deviations or standard errors. The only exception was for

Zhong and Liljenquist's original Study 2, for which the effect size was computed on the basis of the F-test. For the gift choice results, the effect size was computed from 2 x 2 frequency tables. When there were two dependent variables in the same study, the effect sizes were averaged. There are mistakes in the results of both Studies 1 and 3 as published in Zhong and Liljenquist (2006). Therefore, effect sizes were calculated based on personal communication with the first author for Study 1, and a published correction for Study 3. The first author of this meta-analysis and one of the other authors separately extracted and recorded all data used to compute effect sizes to identify any potential discrepancies or errors and thereby ensure accuracy.

Homogeneity. Heterogeneity of the effect sizes was assessed using the Q and I^2 statistics. Heterogeneity among effect sizes is evident from a significant Q statistic (Hedges & Olkin, 1985). I^2 is an index of the extent of heterogeneity, rather than a statistic for a significance test of whether it exists. In particular, I^2 indicates the percent of variance in the set of effect sizes that is accounted for by between-study heterogeneity. As a heuristic, I^2 values of 25 or less, denoting 25% or less of the variability caused by heterogeneity, are understood to reflect low or mild heterogeneity, and values of 50 are understood to reflect moderate heterogeneity (Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). Evidence of heterogeneity indicates the presence of moderating conditions such that the variability in effect sizes across the set of studies is attributed to non-random causes (e.g., methodological and sample differences). When there is no evidence of heterogeneity, the effect sizes across the set of studies are assumed to vary only as a function of sampling error.

Results

The overall weighted mean effect size was small, $g = 0.17$, $p = .013$, 95% CI [0.04 – 0.31], see Figure 1. There was a non-significant trend indicating evidence of mild-to-moderate

heterogeneity, $Q(14) = 23.51$, $p = .053$, $I^2 = 40$. The effect size was little changed when excluding unpublished studies, $g = 0.23$, $p = .003$, 95% CI [0.08 – 0.39], and there was no significant heterogeneity, $Q(11) = 15.98$, $p = .142$, $I^2 = 31$.

The weighted effect size for the three studies first reported in the original publication (Zhong & Liljenquist, 2006) was moderate-to-large, $g = 0.71$, $p < .001$, 95% CI [0.33 – 1.09]. In contrast, the weighted effect size for all subsequent replication efforts from independent laboratories was small and non-significant, $g = 0.07$, $p = .207$, 95% CI [-0.04 – 0.19], and there was no significant heterogeneity, $Q(10) = 11.16$, $p = .345$, $I^2 = 10$.

Discussion

Across 15 studies, the overall effect size for the Macbeth Effect was small. In fact, although the authors of the initial three studies report a moderate-to-large effect, there was no effect whatsoever across 11 subsequent replication attempts by independent laboratories. Therefore, overall, the evidence suggests either that unethical primes do not generate a greater preference for cleansing-related stimuli than do ethical primes, or they generate a small one.

The present analysis does not indicate that there is no relationship between moral and physical hygiene. In fact, Zhong and Liljenquist's finding that participants who cleansed their hands following an unethical prime were subsequently less willing to engage in altruistic behavior than those who had not cleansed - i.e., that cleansing alleviates moral threat whether performed for that purpose or not - has been replicated and extended (Kalanthroff, Aslan & Dar, 2017; Schnall, Benton & Harvey, 2008). We are concerned here exclusively with the reverse, namely whether a moral threat induction is sufficient to motivate cleansing behavior.

While we find little evidence for this hypothesis, a Macbeth Effect may plausibly manifest within certain boundary conditions. Potential situational moderators proposed in the

literature include type of moral threat, i.e. deontological vs. altruistic (D'Olimpio & Mancini, 2014); localization of the cleansing to the "motor modality" (body part or site in the external world) associated with the threat (Lee & Schwarz, 2010); and the presence vs. absence of alternative means of moral self-affirmation, such as punishing another transgressor (Rothschild, Landau, Keefer, & Sullivan, 2015). On the side of individual differences, obsessive-compulsive symptoms involving cleanliness (e.g., Reuven, Liberman, & Dar, 2014) and disgust sensitivity (e.g., Schaefer, in press) have received attention. Broadly speaking, anyone who experiences intense reinforcement of an association between virtue and physical cleanliness would seem more likely than the general population to exhibit a Macbeth Effect. In particular, this is true of many people with a devout religious background.

To researchers interested in the Macbeth Effect, we recommend careful attention to mechanisms through which it may operate. Several existing models seem useful in generating this degree of theoretical precision. The situated inference model, for example, argues that primes increase the accessibility of relevant mental content, which may affect judgment, behavior, or motivation when the origin of that content is misattributed to the self (Loersch & Payne, 2011). In the context of this framework, the strong claim that a particular prime has uniform behavioral effects is implausible. In addition, because the Macbeth Effect is proposed to rely upon a conceptual association between two domains, we refer researchers to relevant work on the psychology of metaphor (e.g., Landau, Meier & Keefer, 2010). In Landau et al.'s framework, metaphor is a tool people use to structure and direct cognitions relating to abstract concepts, such as morality, which may otherwise be experienced as unwieldy or difficult. In fact, the Macbeth Effect has been cited in that literature as a particular case of metaphoric transfer, the mechanism by which cognitions related to a concrete *source concept* (e.g., cleanliness) are

thought directly to impact (by way of *entailments*, or cognitive associations between corresponding features) those pertaining to an abstract *target concept* (e.g., morality) with which they are conceptually paired. The essential point is that metaphoric transfer occurs only in the presence of such specific entailments.

The Macbeth Effect has assumed a high profile since its publication, and in light of the difficulty subsequent researchers have faced in obtaining it, it is important for scholars to know the magnitude of the unmoderated effect. We hope that the present findings spur future research that identifies consistent and replicable moderators, and that this in turn leads to an accurate and nuanced understanding of the phenomenon. Efforts to elucidate boundary conditions begin with the empirical observation that the effect is unreliable, or is much smaller, without identifying those conditions. As psychologists continue to grapple with the reproducibility of our research, we hope that these findings contribute to our communal goal of a replicable psychological science.

Author Note

The authors thank Brian Earp, Chen-Bo Zhong, Francesca D'Olimpio, Jim Everett, and Spike Lee for providing supplemental material; David Wilson for statistical advice; and Victoria Schlaudt and Jamie Ginberg for assistance with manuscript preparation.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*, 217-224. doi:10.1016/j.jesp.2013.10.005

Denke, C., Rotte, M., Heinze, H.-J., & Schaefer, M. (2016). Lying and the subsequent desire for toothpaste: Activity in the somatosensory cortex predicts embodiment of the moral-purity metaphor. *Cerebral Cortex, 26*, 477-484. doi:10.1093/cercor/bhu170

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188.

D'Olimpio, F., & Mancini, F. (2014). Role of deontological guilt in obsessive-compulsive disorder-like checking and washing behavior. *Clinical Psychological Science, 2*, 727-739. doi: 10.1177/2167702614529549

Doyen, S., Klein, O., Simons, D. J., & Cleeremans, A. (2014). On the other side of the mirror: Priming in cognitive and social psychology. *Social Cognition, 32*, 12-32. doi:10.1521/soco.2014.32.supp.12

Earp, B. D. (2017). The need for reporting negative results—a 90 year update. *Journal of Clinical and Translational Research, 3(S2)*, 1-4. doi:10.18053/jctres.03.2017S2.001

* Earp, B. D., Everett, J. A. C., Madva, E. N., & Hamlin, J. K. (2014). Out, damned spot: Can the "Macbeth Effect" be replicated? *Basic and Applied Social Psychology, 36*, 91-98. doi:10.1080/01973533.2013.856792

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in

social psychology. *Frontiers in Psychology, 6*, 621. doi:10.3389/fpsyg.2015.00621

* Fayard, J. V., Bassi, A. K., Bernstein, D. M., & Roberts, B. W. (2009). Is cleanliness next to godliness? Dispelling old wives' tales: Failure to replicate Zhong and Liljenquist (2006). *Journal of Articles in Support of the Null Hypothesis, 6*, 1539-8714.

* Gamez, E., Diaz, J. M., & Marrero, H. (2011). The uncertain universality of the Macbeth Effect with a Spanish sample. *The Spanish Journal of Psychology, 14*, 156-162. doi:10.5209/rev_SJOP.2011.v14.n1.13

Gollwitzer, M., & Melzer, A. (2012). Macbeth and the joystick: Evidence for moral cleansing after playing a violent video game. *Journal of Experimental Social Psychology, 48*, 1356-1360.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*, 193-206.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*, 640-648.

Kalanthroff, E., Aslan, C., & Dar, R. (2017). Washing away your sins will set your mind free: Physical cleansing modulates the effect of threatened morality on executive control. *Cognition and Emotion, 31*, 185-192. doi:10.1080/02699931.2015.1086313

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition.

Psychological Bulletin, 136, 1045-1067. doi:10.1037/a0020970

* Lee, S. W. S., & Schwarz, N. (2010). Dirty hands and dirty mouths: Embodiment of the moral-purity metaphor is specific to the motor modality involved in moral transgression. *Psychological Science, 21*, 1423-1425. doi:10.1177/0956797610382788

Levine, T., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlation in meta-analysis: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs, 76*, 286-302.

Loersch, C. & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science, 6*, 234-252.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*, 615-631. doi: 10.1177/1745691612459058

Protzko, J. & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. In S. O. Lilienfeld & I. W. Waldman (Eds.), *Psychological Science under Scrutiny: Recent Challenges and Proposed Solutions* (pp. 85-107). Chichester: John Wiley & Sons.

Reuven, O., Liberman, N., & Dar, R. (2014). The effect of physical cleaning on threatened morality in individuals with obsessive-compulsive disorder. *Clinical Psychological Science, 2*, 224-229. doi:10.1177/2167702613485565

Rothschild, Z. K., Landau, M. J., Keefer, L. A., & Sullivan, D. (2015). Another's punishment cleanses the self: Evidence for a moral cleansing function of punishing transgressors. *Motivation and Emotion, 39*, 722-741. doi:10.1007/s11031-015-9487-9

* Schaefer, M. (in press). Morality and soap in engineers and social scientists: The Macbeth effect interacts with professions. *Psychological Research*. doi:10.1007/s00426-017-0937-8

Schaefer, M., Rotte, M., Heinze, H.-J., & Denke, C. (2015). Dirty deeds and dirty bodies: Embodiment of the Macbeth effect is mapped topographically onto the somatosensory cortex. *Scientific Reports, 5*, 18051. doi:10.1038/srep18051

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science, 19*, 1219-1222. doi:10.1111/j.1467-9280.2008.02227.x

* Siev, J. (2008). Unpublished raw data.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*, 76-80.

Thomas, J., Grey, I., & Al-Romaithi, S. (in press). Macbeth Arabia: Moral threat and cleansing-related construct accessibility in Arab women. *Mental Health, Religion and Culture.* doi:10.1080/13674676.2017.1381675

* Zhong, C.-B. (2007). Unpublished raw data.

* Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*, 1451-1452. doi:10.1126/science.1130726

Table 1.

Studies Included and Excluded from Meta-Analysis

Study	Manipulation	Outcome	Other Differences from Original Studies
Original Publication:			
Zhong & Liljenquist – Study 1	Recall and describe unethical or ethical deed	Number of cleansing-related solutions generated during word completion task	
Zhong & Liljenquist – Study 2	Copy unethical or ethical story	Desirability of five cleansing-related products	
Zhong & Liljenquist – Study 3	Same as Z&L 1	Choice of gift: antiseptic cleansing wipe or pencil	
Others Included:			
Earp et al. (2014) – Study 1	Same as Z & L Study 2	Same as Z & L Study 2 with British brands substituted for American brands, and with ratings of both desirability and price willing to pay	British participants

Earp et al. (2014) – Study 2	Same as Z & L Study 2	Same as Z & L Study 2 with ratings of both desirability and willingness to pay	Study conducted online
Earp et al. (2014) – Study 3	Same as Z & L Study 2	Same as Z & L Study 2 without brand names provided, and with ratings of both desirability and price willing to pay	Indian participants; Study conducted online
Fayard et al. (2009) – Study 1	Same as Z & L Studies 1 and 3	Same as Z & L Study 3	Participants completed personality measures before manipulation
Gamez et al. (2011) – Study 1	Same as Z & L Studies 1 and 3	Same as Z & L Study 1 with more word fragments	Spanish participants; Materials presented in Spanish
Gamez et al. (2011) – Study 2	Variation on the story used in Z & L Study 2	Same as Z & L Study 2 with ten cleansing-related products instead of five, and brands well-known in Spain	Spanish participants; Materials presented in Spanish; Participants asked about religious beliefs
Gamez et al. (2011) – Study 3	Same as Z & L Studies 1 and 3	Same as Z & L Study 3 with the choice of antiseptic wipe or pen instead of pencil	Spanish participants; Materials presented in Spanish

Lee & Schwarz (2010)	Variation on Z & L Study 2 requiring participants to enact the scenario that participants copied in the original study	Same as Z & L Study 2 with two cleansing-related products instead of five, and with ratings of both desirability and price willing to pay	
Schaefer (in press)	Same as Z & L Study 2	Same as Z & L Study 2 with ten cleansing-related products instead of five	German participants; Materials presented in German; Participants were engineering or social science students to test whether professional domain moderated outcome; Participants completed personality and affect measures before manipulation
Siev (unpublished) – Study 1	Same as Z & L Study 2	Same as Z & L Study 1 with more word fragments and same as Z & L Study 2	Study conducted online
Siev (unpublished) – Study 2	Same as Z & L Study 2	Same as Z & L Study 2	
Zhong (unpublished)	Same as Z & L Study 2	Same as Z & L Study 2 with ratings of both desirability and price willing to pay	

Excluded:

Study	Reason(s) Excluded
Denke et al. (2016)	Within-subjects repeated-measures design. Each participant was presented with 60 scenarios while in an fMRI scanner.
D'Olimpio & Mancini (2014)	Conceptually relevant but not a direct replication effort. The authors presented stories designed to induce guilt and then measured several outcomes including cleaning behavior. Of note, however, the authors specifically created two types of stories designed to induce different types of guilt (deontological versus altruistic). Aside from the overall discrepant methodological approach, it is not clear whether either – and if so, which – of the inductions is consistent with the approach used by Zhong and Liljenquist. The novelty of this study is specifically that the effects of moral challenges on physical cleansing behavior may depend on nuanced considerations related to type of moral emotion.
Earp et al. (2008)	Data were not available to calculate an effect size, although the authors report that the Macbeth Effect did not replicate in this study.
Gollwitzer & Melzer (2012)	Conceptually relevant but not a direct replication effort. The authors had participants play violent video games and then select hygiene or non-hygiene products. The authors reported no main effect of game type on cleansing-related preferences. However, inexperienced (but not experienced) players did select more hygiene products after playing violent games in which the target of the violence was human versus an object. This methodology is very different from Zhong and Liljenquist, and requires the assumption that playing violent video games is the same as recounting or writing about something unethical. Moreover, these authors distinguished between experienced and unexperienced video game players, as well as between violence against humans and objects, neither of which has a parallel in the original studies or replication efforts.

Rothschild et al. (2015)	Participants received a manipulation between the prime and the physical dirtiness ratings. In addition, the authors asked participants explicitly to rate how dirty they felt, thereby providing participants with a cleanliness metaphor. In fact, participants may have simply understood the questions about cleanliness metaphorically. In contrast, the dependent variables in included studies were implicit or not recognized (and often intentionally hidden).
Schaefer et al. (2015)	Within-subjects repeated-measures design. Each participant was presented with 60 scenarios while in an fMRI scanner.
Thomas et al. (in press)	The control group did not receive any prime. In addition, the prime in the experimental group was a newly developed four-item questionnaire that is not comparable to any priming task used in the original studies or replication efforts.

Note. Z & L = Zhong & Liljenquist

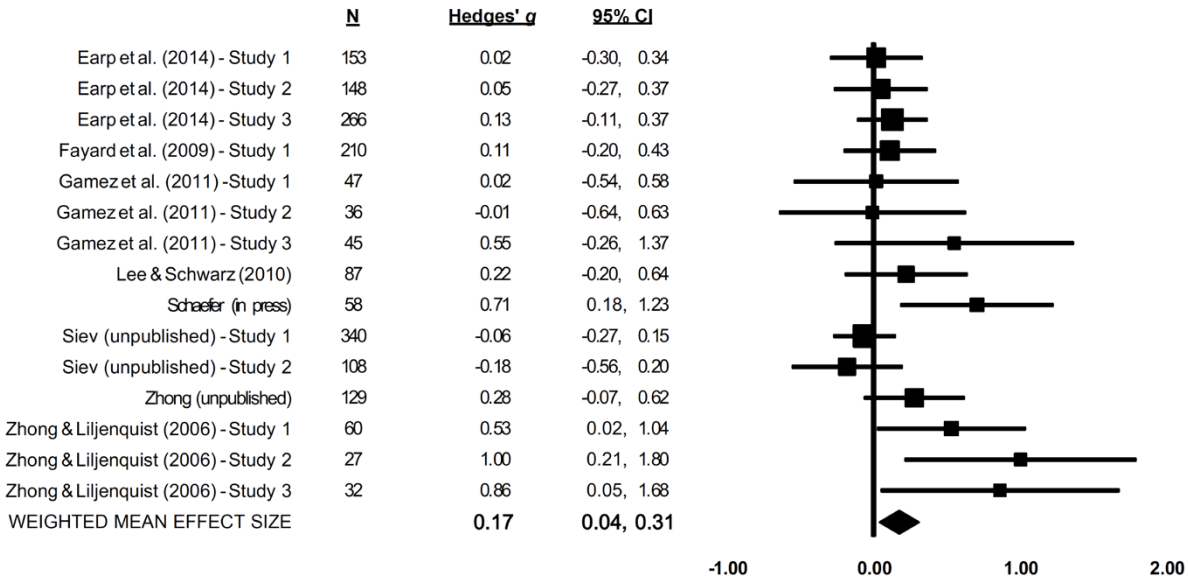

Figure 1. Forest plot for random-effects meta-analysis of the effect of unethical primes on cleansing preference. Dot size is proportional to study weight. Error bars represent 95% confidence intervals (CI). Positive effect sizes denote greater cleansing preference for the unethical condition than the ethical condition.