(in press, *Psychological Bulletin*)

Wipe It Off: A Meta-Analytic Review of the Psychological Consequences and Antecedents of Physical Cleansing

Spike W. S. Lee, Kathleen Chen, Cecilia Ma, Joe Hoang
University of Toronto

Author note

All of our data, code, and research materials (including our coding scheme) have been made publicly available at the Open Science Framework and can be accessed at https://osf.io/xv2wj. This review was not preregistered, but we have strived to satisfy all other relevant criteria on the AMSTAR 2 inventory (Shea et al., 2017). We also adhere to all PRISMA (Moher et al., 2009) and MARS (American Psychological Association, 2020; Appelbaum et al., 2018, pp. 21–23) guidelines for meta-analytic reporting. Our self-assessments using both the AMSTAR 2 inventory and the PRISMA checklist are provided in the supplemental material.

Lee designed the meta-analysis, coded moderators, conducted statistical analyses, wrote the manuscript, and revised it. Chen coded moderators, extracted data, and computed effect sizes. Ma and Hoang extracted study information and data for supplemental exploratory analyses.

This work was supported by the Early Researcher Award (Round 13, Ontario Ministry of Research, Innovation and Science), Insight Grant (435-2017-0127, Social Sciences and Humanities Research Council, SSHRC), Insight Development Grants (430-2021-00060 and 430-2015-00333, SSHRC), and New Researcher Award (2014-2015, Connaught Fund) awarded to Spike W. S. Lee; the University of Toronto Excellence Award (2018) awarded to Kathleen Chen; the Joseph-Armand Bombardier Canada Graduate Scholarship-Doctoral (2021-2025, SSHRC) awarded to Cecilia Ma; and the Canada Graduate Scholarship-Master's (2019-2020, SSHRC) awarded to Joe Hoang. None of the funders was involved in conducting the present meta-analysis or deciding to publish the results. We declare no conflict of interest.

We thank Eileen Zheng Wu, Ethan Li, Frankline George, Joanne Lu, and Thalia Vrantsidis for research assistance; Ivan Ropovik, Alessandro Sparacio, Hans IJzerman, Robert M. Ross, Robbie C. M. van Aert, Olmo R. van den Akker, Michiel van Elk, and Joseph Hilgard for suggestions of bias assessment and correction methods; Julie Y. Huang, Norbert Schwarz, Phillip R. Johnston, action editor Sheri Johnson, Editor Blair Johnson, and three reviewers for feedback.

Spike W. S. Lee, Rotman School of Management and Department of Psychology, University of Toronto; Kathleen Chen, Department of Economics and Department of Mathematics, University of Toronto; Cecilia Ma and Joe Hoang, Department of Psychology, University of Toronto.

Correspondence concerning this article should be addressed to Spike W. S. Lee, Rotman School of Management and Department of Psychology, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6, Canada. Email: spike.lee@utoronto.ca

Abstract

Physical cleansing is a human universal. It serves health and survival functions. It also carries rich psychological meanings that interest scholars across disciplines. What psychological effects result from cleansing? What psychological states trigger cleansing?

The present meta-analysis takes stock of all experimental studies examining the psychological consequences and antecedents of cleansing-related thoughts, feelings, and behaviors (e.g., feeling less guilty after cleansing; spontaneously cleansing oneself after thinking of unwelcomed sexual encounter). It includes 129 records, 230 experiments, and 551 effects from 42,793 participants. Effect sizes were synthesized in random-effects models using robust variance estimates with small-sample corrections, supplemented by other techniques. Outliers were excluded using leave-one-out diagnostics and sensitivity analysis. Publication bias was assessed and corrected for using eight methods. Theoretical, methodological, sample, and report moderators were coded.

After excluding outliers, without bias correction, the synthesized effect size estimate was g = 0.315, 95% CI [0.277, 0.354]. Using various bias correction methods, the estimate ranged from g = 0.103 to 0.331 and always exhibited considerable heterogeneity. Effect sizes were especially large for behavioral measures and varied significantly between sample types, sample regions, and report types. Meanwhile, effects were domain-general (observed in the moral domain and beyond), bidirectional (physical cleansing ↔ psychological variables), and robust across theoretical types, manipulation operationalizations, and study designs. Limitations included mixed replicability, suboptimal methodological rigor, and restricted sample diversity. We recommend future studies to (a) incorporate power analysis, preregistration, and replication, (b) investigate generalizability across samples, (c) strengthen discriminant validity, and (d) test competing theoretical accounts.

Public Significance Statement

Cleansing behavior happens throughout the day. Not only does it benefit personal and public health, but it also carries rich psychological and cultural meanings. Our meta-analytic review synthesizes causal evidence from 551 effect sizes and reveals robust effects of manipulating physical cleansing (e.g., actual or imagined handwashing) on psychological outcomes across domains (e.g., morality, politics, religion, emotion, motivation, judgment and decision making). It also reveals robust effects of manipulating diverse psychological experiences (e.g., presence of others, imagined unwelcomed sexual encounter) on cleansing-related thoughts, feelings, and behaviors. All together, our findings highlight the psychological power and social ramifications of physical cleansing.

Keywords

cleansing, morality, conceptual metaphor, grounded procedure, meta-analysis

Wipe It Off: A Meta-Analytic Review of the Psychological Consequences and Antecedents of Physical Cleansing

Physical cleansing is part of our everyday life. It takes many forms. It can involve cleansing oneself, such as hand-washing, face-cleaning, teeth-brushing, mouth-rinsing, nail-clipping, shaving, and showering. Or it can involve cleansing one's surroundings, such as wiping the table, doing the dishes, clearing the garbage, mopping the floor, vacuuming the carpet, and doing the laundry. In addition to these mundane routines, more elaborate forms of sanitizing and disinfecting may be normalized under special circumstances (e.g., in hospital wards, during a contagious disease pandemic).

Indeed, year after year, data from the American Time Use Survey (ATUS) attest to the pervasiveness of cleansing behavior. Take the latest ATUS prior to the COVID-19 pandemic[1] (U.S. Bureau of Labor Statistics, 2019) as an example. Across all categories and subcategories of activities, grooming was the fourth highest in terms of the average percentage of the civilian population who engaged in relevant activities per day (80.7%), only after necessities such as sleeping (100.0%), eating and drinking (95.8%), and socializing, relaxing, and leisure (94.1%). Examples of grooming, as defined by the ATUS, ranged from "washing face," "washing hands," and "shaving legs" to "bathing/showering," "brushing/flossing teeth," and "brushing lint off clothing." Among respondents who engaged in grooming activities, women spent a daily average of 57 minutes, and men 43 minutes. Following grooming, the next major category was household activities (78.3%), with examples such as interior cleaning (22.6%; daily average 81 minutes among respondents who engaged in it), laundry (16.6%; 64 minutes), and kitchen and food cleanup (23.0%; 33 minutes). Whether we like it or not, cleansing behaviors constitute a non-trivial part of our daily time use.

Why are people willing to put their time and effort into such uninspiring activities? One answer could be that, like many things in life, people do it not because they want to, but because they have to. Social norms require them to keep themselves, their possessions, and their surroundings clean. Although normative expectations about the exact form and frequency of cleansing show some variation across cultures and history (Ashenburg, 2007; Hoy, 1995), the need for and presence of hygienic care is a human universal (Brown, 1991). It makes biological sense, insofar as hygiene confers personal and public health benefits and increases chances of survival. Handwashing, for example, is among the cheapest, easiest, and most effective mechanisms for reducing a wide array of disease risks (Boyce & Pittet, 2002; Kampf & Kramer, 2004). It is one of the most recommended routines by the World Health Organization (Pittet et al., 2009). Awareness about its affordability, effectiveness, and best practices is raised every year on October 15, the Global Handwashing Day.

But besides health, other factors may also underlie the regular occurrence of cleansing behavior. The past 15 years have witnessed a growing body of evidence for the diverse psychological consequences and antecedents of physical cleansing, which will be collectively referred to as "cleansing effects" throughout this paper. Cleansing manipulations have been shown to reduce the influence of a variety of recent experiences, as if mentally "wiping the slate clean." For example, wiping or washing one's hands can reduce the influence of immoral recall on one's negative emotion and guilt-compensatory volunteering behavior (Zhong & Liljenquist, 2006), reduce the influence of a product choice on one's subsequent product evaluation (S. W. S.

[1] We report the latest ATUS data prior to the COVID-19 pandemic because public health guidelines have heightened attention to cleansing behaviors during the pandemic, rendering their estimated frequencies atypical and unrepresentative of normal times. Also, according to the U.S. Bureau of Labor (2022), "The 2020 ATUS was greatly affected by the coronavirus (COVID-19) pandemic. Data collection was suspended in 2020 from mid-March to mid-May for the safety of ATUS staff. Annual 2020 estimates cannot be produced due to the 2-month suspension in data collection, and thus ATUS tables were not updated with 2020 data."

Lee & Schwarz, 2010b; also De Los Reyes et al., 2012), reduce the influence of an academic failure on one's pessimism about one's future performance (Kaspar, 2012), reduce the influence of a product endowment on one's subsequent attachment to it (Florack et al., 2014), reduce the influence of a lucky or unlucky streak on one's subsequent betting behavior (A. J. Xu et al., 2012; also Moscatiello & Nagel, 2014), and more. Conversely, cleansing-related thoughts, feelings, and behaviors can be elicited by imagining an unpleasant sexual encounter (Fairbrother et al., 2005), recalling one's immoral behavior (Zhong & Liljenquist, 2006), processing religion-related concepts (Preston & Ritter, 2012), smelling a shirt belonging to an outgroup member (Reicher et al., 2016), thinking about ways to change one's life (Jiang & Gao, 2015), and more. This body of research, instead of using correlational or observational approaches common in earlier work, relied on experimental methods to examine causal relations between cleansing and many basic psychological aspects of daily life, such as morality, politics, religion, emotion, motivation, and judgment and decision making. A precise understanding of these causal links entails examining the robustness, scope, nuances, and mechanisms of cleansing effects. We do so quantitatively through a meta-analysis.

Theoretical Interest in Cleansing Effects

Cleansing is a topic of significance to various disciplines and with diverse manifestations in the wild. For example, anthropological work (Douglas, 1966) has long pointed out that cleansing and purity are imbued with symbolic meanings such as threat and danger, not only in "primitive cultures" (p. 74) or religious communities, but also in modern and secular ones. Religious scholars have noted the prevalence of purification rituals across all major religions in the world, involving billions of people. From baptism in Christianity and achamanam in Hinduism to corpse-rinsing before burial in ancient Egypt, physical cleansing is construed as a way to renew the body, soul, and spirit (Blackman, 1918; Eliade, 1958/1996; Michael, 1979). An English theologian in the 18th century, likely inspired by a bountiful supply of relevant verses throughout the Old and New Testaments, put cleanliness right next to godliness (Wesley, 1778). And pastors in the 21st century have designed a 35-day plan called Soul Detox (Life.Church, n.d.) to "help you identify what's chipping away at your soul, and what's getting in the way of you becoming the person who God created you to be. You will learn from God's Word how you can neutralize these damaging influences and embrace clean living for your soul."

These examples illustrate the power of subjective construal (Griffin & Ross, 1991; Taylor, 1998) in imbuing a physical act of cleansing with layers of socially shared and subjectively felt psychological meanings. A particularly astonishing case is the Ganga River in Allahabad, India. One of the most polluted rivers on earth (The World Bank, 2011; Zerkel, 2013), it is also "one of the holiest spots in Hinduism. Allahabad, Persian for 'Settled by God,' plays host every dozen years to the Kumbh Mela, the biggest gathering of humanity on Earth, when tens of millions of pilgrims come to wash away their sins" (Morrison, 2011). Washing oneself in a physically dirty river, through a socially constructed interpretation, can be morally and spiritually cleansing.

The naturalistic prevalence and surprising power of physical cleansing have piqued psychologists' interest. Our goal in this meta-analytic review, put simply, is to unpack the mental underpinnings of cleansing effects. We do so by integrating and juxtaposing data from multiple lines of experimental research. We ask a straightforward but ambitious research question: What are all the psychological consequences and antecedents of physical cleansing that have been documented in the literature? Synthesizing the evidence will allow us to fully depict the empirical landscape, which in turn will inform theorizing about the properties and functions of cleansing in multiple areas of psychology.

Specifically, work in cognitive linguistics has associated concrete experiences of cleanliness with abstract thoughts about morality (Lakoff & Johnson, 1980, 1999). This "clean–moral" association (S. W. S. Lee & Schwarz, 2016) is linked to neural activities in modality-

specific sensorimotor regions of the brain (Denke et al., 2014; Schaefer et al., 2015; Tang et al., 2017), compatible with the broader perspective of grounded cognition that assumes all forms of cognition to be grounded in sensorimotor modalities (Barsalou, 1999, 2008). These observations about the association of cleanliness with morality lead to the empirical expectation that cleansing effects should occur primarily in the moral domain.

Meanwhile, work in affective science and social psychology has identified cleansing as an effective behavior for reducing disgust and thus reducing the effects of disgust in moral and related domains such as contagion, prejudice, and religion (J. Y. Huang et al., 2011; Nemeroff & Rozin, 1994; Ottaviani et al., 2018; Ritter & Preston, 2011; Schnall et al., 2008; Zhong et al., 2010; Zhong & Liljenquist, 2006). Conversely, disgust and processes in which disgust plays a role have been found to elicit thoughts about, desires for, and actions of cleansing (Galoni & Noseworthy, 2015; Golec de Zavala et al., 2014; S. W. S. Lee et al., 2015; S. W. S. Lee & Schwarz, 2010a). These observations point to the link between cleansing and disgust, which is not restricted to the moral domain.

Beyond morality and disgust, studies of animal behavior have shown that stressors such as loud noises and conspecifics' screams elicit self-cleansing behavior among many species (e.g., pine voles, rats, primates; Leung & Borst, 1987; Mason et al., 1985; Reidinger et al., 1982; Spruijt et al., 1992). Dovetailing the link between stress and cleansing in animals, recent human experiments have found that cleansing reduces the affective and physiological effects of stressful events (S. W. S. Lee et al., 2023), regardless of whether people actually engage in cleansing behavior or merely simulate it in their mind. Importantly, none of these stressors has anything to do with morality or disgust, suggesting that cleansing effects may be broader than predicted by earlier theories.

To inform these theoretical perspectives, delineating the scope and complexity of cleansing effects will be one of our primary research goals, as elaborated in the next section (**Research Questions**). Another goal will be probing for potential differences between psychological consequences and antecedents of cleansing (i.e., what types of outcomes follow from cleansing and what types of factors result in cleansing?), an issue that bears on competing theoretical predictions, as some theories predict only psychological consequences of cleansing but are silent about psychological antecedents of cleansing, whereas other theories predict both psychological consequences and antecedents of cleansing. An additional cluster of research questions will revolve around methodological issues (e.g., experimental design, replicability), which will enable us to critically evaluate the strengths and weaknesses in the existing body of cleansing effects.

In addition to theoretical and methodological implications, identifying the empirical properties of cleansing effects will have applied implications in health-related contexts. For example, the most common symptoms of obsessive-compulsive disorder (OCD) include obsessive fear of germs or contamination and compulsive behaviors of excessive cleaning or handwashing (National Institute of Mental Health, n.d.; Stanford Medicine, 2022). Discerning the most powerful elicitors of cleansing behavior may shed light on the underpinnings of OCD. Health advocates may also leverage the most robust antecedents of cleansing to promote hygienic behavior. In short, the potential to bring about empirical clarity, with implications for researchers and practitioners alike, motivates us to conduct a comprehensive meta-analytic review of cleansing effects.

Research Questions

Scope and Complexity of Cleansing Effects

Different theoretical perspectives predict different scopes and complexities of cleansing effects. From an anthropological perspective, religious purification rituals are observable around the world and throughout history (Blackman, 1918; Douglas, 1966; Eliade, 1958/1996; Michael,

1979). They reflect a widely shared mental association between purity and divinity (Preston & Ritter, 2012). If religion was the only source of the association, cleansing effects should be specific to religiously relevant content (e.g., sin, sanctification), but not beyond.

From a cognitive-linguistic perspective, however, purity has a broader meaning: "substances that are pure are typically clean" and "when morality is conceptualized as purity and purity as cleanliness, we get the derived metaphor Morality Is Cleanliness"[2] (Lakoff & Johnson, 1999, p. 307). Although some of our moral values do find their roots in religious teaching, not all of them do. That means the conceptual-metaphorical association of purity/cleanliness with morality is different in scope and origin from its anthropological association with religion. To date, the only conceptual metaphors that have been postulated as relevant to cleansing are "Morality Is Cleanliness" and "Morality Is Purity" (Lakoff & Johnson, 1980, 1999). As such, the cognitive-linguistic perspective of conceptual metaphor predicts cleansing effects within the moral domain, but not beyond.

Within moral psychology, scholars differ in their views on the nature of morality, from which different predictions about cleansing effects may be derived. A prominent perspective assumes the existence of multiple foundational moral intuitions (Graham et al., 2009, 2011), including one that is most strongly and directly shaped by concerns with disgust and contamination. This moral foundation, called "sanctity/degradation," underlies religious pursuits of sanctification as well as non-religious pursuits of nobility in character. From this perspective, it may be expected that cleansing effects are restricted to, or at least strongest for, moral violations that evoke disgust reactions and sanctity or nobility concerns such as sexual deviancy (Helzer & Pizarro, 2011; Rozin et al., 2008; Rozin & Fallon, 1987). This perspective thus entails the prediction that within the moral domain, cleansing effects are moderated by moral foundation and mediated by disgust.

In contrast, another perspective assumes that a single dyadic template assessing interpersonal harm underlies all moral judgments (Gray et al., 2012; Schein & Gray, 2018), even for moral violations that on the surface seem not to involve any harm. This perspective does not postulate that harm in any particular subdomain of moral violations has a uniquely strong link to contamination or cleanliness. It thus entails the prediction that cleansing effects are similarly applicable and strong across different subdomains of moral judgment, regardless of whether they evoke disgust (or other emotions; Cameron et al., 2015). That is, it predicts that within the moral domain, cleansing effects are neither moderated by moral foundation nor mediated by disgust; instead, they are observed across moral foundations and mediated by harm.

Although these two perspectives (moral foundations and dyadic morality) make different specific assumptions, both predict cleansing effects only within the moral domain, not beyond. Contrary to them, an emerging perspective assumes that physical cleansing provides experiential grounding for the mental procedure of psychological separation (S. W. S. Lee & Schwarz, 2021). Inspired by the broad notion of grounded cognition (Barsalou, 1999, 2008), this perspective highlights the fact that cleansing behavior involves separating physical traces from a physical target (e.g., removing dirt from one's hands, removing grease off a plate), a motor procedure that can ground the mental procedure of separating psychological traces from a psychological target (e.g., dissociating a prior experience from one's self, dissociating one event from the next). A mental procedure, by definition, is not tied to any particular content; instead, it is applicable across domains (Wyer et al., 2012). From this process-oriented perspective,

[2] The notation of "Morality Is Cleanliness" does not mean that morality is identical to cleanliness. Instead, it signifies a conceptual metaphor wherein conceptual understanding of morality, a rather abstract and complex psychological domain, is structured by conceptual understanding of cleanliness, a more concrete and easier-to-comprehend physical domain. For further elaboration, see the section **Directionality of Cleansing Effects**.

cleansing effects are expected to be domain-general (i.e., not moderated by domain) and mediated by separation.[3]

Motivated by these competing perspectives, the present meta-analysis will address the following empirical questions by examining moderator variables between effects and between studies:

- Are cleansing effects stronger within the moral domain than beyond? (Being stronger can mean various things, such as "observed only here but not there," "larger in size here than there," or "more robust here than there.")
- Within the moral domain, are cleansing effects stronger in cases that evoke sanctity/degradation concerns (which tend to be linked to disgust) than in cases that do not evoke these concerns?
- Among cases that evoke sanctity/degradation concerns, are cleansing effects stronger for sexual content (which tends to be particularly strongly linked to disgust) than for non-sexual content?
- Beyond the moral domain, are cleansing effects stronger in some domains than in others?
- How much of the experimental work provided evidence of moderation within the original study? How much of it provided evidence of mediation?

Directionality of Cleansing Effects

Different perspectives also make different predictions about the directionality of cleansing effects. Some predict that physical cleansing should influence psychological variables (i.e., psychological consequences of cleansing) but are silent about the influence of psychological variables on cleansing-related thoughts, feelings, and behaviors (i.e., psychological antecedents of cleansing). Others predict bidirectional influence between physical cleansing and psychological variables (i.e., psychological consequences and antecedents of cleansing).

Specifically, the perspective of metaphorical structuring or scaffolding (Boroditsky, 2000; Thibodeau et al., 2017; Williams et al., 2009) has received empirical support from behavioral experiments. The basic assumption is that "concepts and goal structures specialized for interacting with the physical environment (e.g., distance cues, temperature, cleanliness, and self-protection), which emerge early and automatically as a natural part of human development and evolution… serve as the foundation for the later development of more abstract concepts and goals" (Williams et al., 2009, p. 1257). Mappings from physical to psychological experiences constitute cognitive representations that are often reflected in the linguistic metaphors that speakers utter frequently, effortlessly, and systematically (Gibbs, 1994; Lakoff & Johnson, 1999). From this perspective, physical cleansing—a basic behavior that is functionally adaptive and that emerges relatively early ontogenetically and phylogenetically—should support higher mental processes and more abstract or complex cognitive representations. It implies a unidirectional effect from the physical to the psychological, or at least an asymmetric effect that is stronger from the physical to the psychological than vice versa (see also IJzerman & Koole, 2011; Landau et al., 2011; Schneider et al., 2011).

Alternative predictions can be derived from the related but distinct perspective of grounded cognition (Barsalou, 2008). It assumes that "modal simulations, bodily states, and

[3] Earlier, we mentioned the power of subjective construal in imbuing a physical act of cleansing with socially shared and subjectively felt psychological meanings. Subjective construal is also involved in experiencing a physical act as separating physical traces in the first place. Consider the example of applying hand sanitizer to one's hands. At first glance, putting something on one's hands does not seem like an example of separating anything from oneself. But to the extent that people construe the application of hand sanitizer as eliminating germs from one's hands, it is experienced as an act of separating physical traces from oneself.

situated action underlie cognition" (p. 617). A key aspect of grounding (Barsalou, 2016) is the principle of neural reuse (Anderson, 2010), which states that neural circuitry for cognitive functions that emerge earlier in evolution or development tends to be exapted for cognitive functions that emerge later. By implication, the neural underpinnings and sensorimotor states of physical cleansing should ground or underlie cognition—that is, they are not tangential to or independent of mental activity; they *constitute* mental activity. Sensorimotor states and cognitive states are tied together. If so, manipulating physical cleansing should exert psychological effects, and vice versa. This predicts bidirectional and symmetric effects between physical cleansing and psychological variables.

To inform these perspectives, the present meta-analysis will ask:

- Across all experiments, are the psychological consequences of physical cleansing comparable to the psychological antecedents of physical cleansing in terms of effect size and heterogeneity?
- Within each of the other moderator variables we examine, are the psychological consequences of physical cleansing comparable to the psychological antecedents of physical cleansing in terms of effect size and heterogeneity? For example, are the psychological consequences of cleansing similar in strength to the psychological antecedents of cleansing in the moral domain, but stronger than the psychological antecedents of cleansing in other domains? Empirical nuances of this sort will be probed by double-moderator analyses that test the interaction between directionality and other moderators of cleansing effects.

Methodological Issues about Cleansing Effects

Several clusters of questions pertaining to methodological properties of cleansing effects will also be addressed by moderator analyses. In terms of experimental operationalizations:

- What are the most effective methods of manipulating or measuring physical cleansing? Is it actual cleansing, or mentally simulated cleansing, or merely activating the concept of cleansing, or are they equally effective?
- What categories of psychological variables exhibit the strongest cleansing effects? Is it judgment, or behavior, or others?

In terms of robustness, failed replications of cleansing effects have been reported (Camerer et al., 2018; Earp et al., 2014; Fayard et al., 2009; Gámez et al., 2011; D. J. Johnson et al., 2014a; Klein et al., 2018; Siev, 2008; Zhong, 2007). Successful demonstrations of cleansing effects have also been reported (as detailed in the present meta-analysis and summarized in prior narrative reviews, e.g., S. W. S. Lee & Schwarz, 2011, 2016; Y. Wang et al., 2019; West & Zhong, 2015; Yan, 2011). Multiple interpretations are plausible. One is that the original effects were statistical flukes. Another is that the original effects were true phenomena limited to specific manipulations, measures, settings, or populations, i.e., they had limited generalizability. Yet another interpretation requires us to zoom out, situate both the original experiments and the replications in the broader context of all relevant effects, evaluate the strength of evidence overall, and probe for variables that track which effects are more replicable and robust than others. For example:

- Is the discrepancy between failed and successful replications attributable to methodological differences between experiments? Or publication bias?
- If the latter, how severe is the problem? Do more recent publications, with more contemporary research practices, tend to report weaker effects? Or stronger effects?

Finally, moderator analyses will also explore methodological issues in terms of generalizability:

- Have cleansing effects been mostly documented in "WEIRD" samples (Western, Educated, Industrialized, Rich, and Democratic; Henrich et al., 2010) such as undergraduate students and online participants in Western societies? Or are cleansing effects observable in other communities as well?

- If so, do the effect sizes vary as a function of sample type and region?

Overview of the Present Meta-Analysis

To provide empirical answers to all of these research questions, the present meta-analysis synthesized causal evidence from experimental research on physical cleansing to estimate the effects on its psychological consequences and the effects of its psychological antecedents. State-of-the-art techniques (e.g., parameter selection model) were used alongside traditional ones (e.g., normal-quantile plot with Shapiro-Wilk normality test) to quantify the evidential value of the meta-analytic dataset and the extent of publication bias. Moderator analyses were structured in such a way as to inform theoretical perspectives and ascertain robustness of effects across operationalizations of manipulation and measure, features of experimental and statistical design, participant demographics, and report characteristics.

Studies were included if they (a) were experimental and (b) involved either of the following: manipulating physical cleansing and measuring other psychological outcomes or manipulating other psychological variables and measuring outcomes related to physical cleansing. Non-experimental results (e.g., correlational studies, archival analyses) were excluded. Experiments that did not involve physical cleansing (e.g., experiments that only involved disgust but not cleansing) were excluded. Experiments that involved physical cleansing as both independent and dependent variables (e.g., experiments that exposed participants to a clean scent and measured cleansing-related thoughts and behaviors; Holland et al., 2005) were excluded.

The experimental focus, empirical scope, and analytic complexity of the present meta-analysis render it unique from all relevant prior reviews. Specifically, (1) a recent meta-analysis examined "the effects of an immoral versus moral prime on cleansing-related preferences or behaviors" (Siev et al., 2018, p. 2). It included 15 effect sizes, capturing a subset of our dataset. Ours includes (a) their effects, (b) effects of psychological variables other than morality on any cleansing-related variable, and (c) effects of cleansing on any psychological variable, totalling 551 effect sizes. Their meta-analysis included no assessment of publication bias or moderators. Ours does.

(2) A few narrative reviews of cleansing effects are available, including two brief and dated ones (S. W. S. Lee & Schwarz, 2011, 2016), a recent one that added more recent cleansing effects but only selectively reviewed illustrative findings (S. W. S. Lee & Schwarz, 2021), one that specifically examined cleansing effects in relation to morality (West & Zhong, 2015), and one with a similar goal that was published in Chinese (Yan, 2011). None of these was comprehensive or quantitative.

(3) A related construct that has received much scientific attention in the last decade is disgust, especially in relation to morality (for related reviews, see Cameron et al., 2015; Chapman & Anderson, 2013; Feder, 2016; Giner-Sorolla & Sabo, 2016; Horberg et al., 2011; S. W. S. Lee & Ellsworth, 2013; Olivera La Rosa & Rosselló Mir, 2013; Pizarro et al., 2011; Pole, 2013; Rozin et al., 2008; Russell & Giner-Sorolla, 2013; Schnall, 2014, 2017; Strohminger, 2014; Tybur et al., 2013; Widen & Russell, 2013). A meta-analysis included 50 effects of incidental disgust and moral judgment (Landy & Goodwin, 2015).

Although relevant to our interest, disgust is distinct from cleansing in a couple of ways. (a) Disgust overlaps with dirty (the opposite of clean) partially, but not completely. Many dirty things are disgusting (e.g., feces in a toilet bowl), but some dirty things are not disgusting (e.g., dust on a TV screen) and some disgusting things are not dirty (e.g., body disfigurement). (b) Even if we focus on the overlapping area, cleansing is positive, whereas disgust is negative. Manifestations of one may not be reducible to manifestations of the other, because each valenced experience may have its own unique properties. Much as a comprehensive analysis of pleasure cannot be reduced to that of the absence of pain, a comprehensive analysis of cleansing cannot be reduced to that of the absence of dirtiness and disgust (Schnall, 2011). Indeed,

cleansing effects extend far beyond disgusting contexts, as will be seen in our meta-analytic results.

Method

This section describes the methodological details of identifying relevant experiments and effects, computing and synthesizing effect sizes, assessing outliers and publication bias, and examining theoretical, methodological, participant, and report moderators. Figure 1 depicts the overall process in a flowchart. The final meta-analytic dataset consisted of 129 unique records, 230 experiments, and 551 effects based on 42,793 participants from different world regions. Full references for all meta-analyzed records are available both in the References list and in the supplemental material on the Open Science Framework (OSF) at https://osf.io/xv2wj/.

Literature Search

To identify relevant research in the published and gray literature across disciplines (e.g., psychology, consumer behavior) and languages (e.g., English, German) from diverse sources, we searched each of the following:

(1) APA PsycInfo, which covers journal articles, books, book chapters, and dissertations in psychology and related items in cognate fields (e.g., sociology, political science, business, law, education, neuroscience, biology), including journals from 29 languages, with comprehensive coverage starting in the 1880s;
(2) all digitally available conference proceedings of the Society for Personality and Social Psychology (2003-2022), Society of Experimental Social Psychology (2011-2021), Association for Psychological Science (2013-2021; also 2015 & 2017 & 2019 International), Society for Consumer Psychology (2005-2022), and Association for Consumer Research (1971-2021, including its North American Advances, Asia-Pacific Advances, European Advances, Gender and Consumer Behavior, Latin American Advances, and Special Volumes);
(3) PsychFileDrawer.Org; and
(4) all recent large-scale replication projects, including the Reproducibility Project: Psychology, Social Sciences Replication Project, and Many Labs 1, 2, 3, 4, and 5.
(5) We also requested unpublished data via the listservs of the Society for Personality and Social Psychology, Society of Experimental Social Psychology, and Society for Consumer Psychology.

Our search in (1) APA PsycInfo (provided by ProQuest) used the following criteria and concluded at the end of February 2019, rendering 11,346 items:

- Function: “Advanced Search”
- Keywords in “Anywhere”: (cleanse OR cleanses OR cleansing OR cleansed OR clean OR cleans OR cleaning OR cleaned OR cleanliness OR clean*) OR (wipe OR wipes OR wiping OR wiped OR wash OR washing OR washed OR wash*) OR (purify OR purifying OR purified OR purification OR pure OR purity)
- “Methodology”: “Quantitative Study”
- “Population”: “Human”
- “Results page options”: “Exclude duplicate documents” and “Show additional terms included in the search”

The keywords in the second bullet point were also used for our search in (2) conference proceedings (rendering 73 items), (3) PsychFileDrawer.Org (rendering 5 items), and (4) replication projects (rendering 3 items). (5) Our request for unpublished data via scholarly listservs was based on the requirement that the study included a manipulation or measure

related to physical cleansing, cleanliness, purification, or purity (rendering 3 items). These searches concluded at the end of April 2022.[4]

For each rendered item, we examined its title and abstract to decide its relevance (see **Inclusion Criteria**). If uncertain, we read its full-text or examined the variables in its dataset to decide. Items written in languages other than English (e.g., Chinese) were either comprehended by native speakers on our research team or translated by other native speakers for our research team to read. Each item was examined by two authors[5] for decision of inclusion vs. exclusion; any disagreement was noted and resolved by discussing the item at hand in conjunction with other included and excluded items to ensure consistency.

Inclusion Criteria

For a research study to be included in the present meta-analysis of experimental evidence for cleansing effects, it had to meet all of the following criteria:

(1) It had to report primary empirical data. Therefore, secondary analyses of archival data, meta-analyses of empirical data, and literature reviews with no data were excluded.
(2) It had to report quantitative data. Therefore, purely qualitative research was excluded.
(3) It had to involve human participants. Therefore, animal research was excluded.
(4) It had to be a true experiment, i.e., it had to involve random assignment of participants into manipulated conditions, which could be between-participant or within-participant. Therefore, purely correlational studies (whether cross-sectional or longitudinal) were excluded.
(5) The experiment had to *either* (a) manipulate physical cleansing and measure other psychological variables *or* (b) manipulate other psychological variables and measure outcomes directly about physical cleansing. Therefore, experiments that did not

[4] Some of the reports that were available before the end of April 2022 and included in our meta-analysis (see **Inclusion Criteria**) were going through the publication process while we were working on this paper. For example, during our revision process, an unpublished report and a conference presentation (which had already been available before the end of April 2022) were published as peer-reviewed journal articles. In such cases, we cite their most up-to-date version (e.g., S. W. S. Lee et al., 2023; Meng & Gamlin, 2023) and use it for the purpose of our moderator coding (e.g., their "report type" would be coded as journal article, not unpublished report or conference presentation).

[5] Examination was conducted by the first author together with either the second or third author. Following APA's meta-analysis reporting standards (Appelbaum et al., 2018), we describe the "[q]ualifications (e.g., training, educational or professional status) of those who conducted each step in the study selection process" and "the data extraction process" (p. 22). The first author (PhD in social psychology, tenured faculty member) had prior firsthand experience with planning, conducting, and publishing a meta-analysis from start to finish. For study selection, the first author trained the second author (third-year undergraduate student at the time, currently researcher at an economics consulting firm) and third author (post-baccalaureate at the time, currently graduate student in social psychology) on the comprehensive goals of the present meta-analysis, the detailed criteria for study inclusion and exclusion, and the importance of maintaining consistency across studies. For data extraction, the first author trained the second author to understand and use the statistical methods and computer programs required for computing effect sizes from reported information, and to code all moderators, until they were able to execute these tasks on their own. The first author also trained the third author and fourth author (graduate student in social psychology) to extract study information and data for supplemental exploratory analyses. Ongoing supervision and iterative discussion guided the processes of study selection, data extraction, and moderator coding. Inter-rater reliability was not computed because throughout the coding process, any disagreement was noted and resolved by discussion. A rough estimate would be 90% overall, given the high level of inter-rater agreement on the coding of all moderators prior to any discussion.

involve physical cleansing were excluded (e.g., experiments that only involved disgust but not cleansing). Experiments that involved only physical cleansing but no other psychological variable were also excluded (e.g., experiments that exposed participants to a clean scent and measured thoughts about and behaviors of cleansing; Holland et al., 2005).

For the same conceptual reason, if an experiment manipulated physical cleansing and measured multiple outcomes, some of which were directly about cleansing and some were not, then in order to examine links of cleansing to *other* psychological variables, only effects pertaining to outcomes not about cleansing were included. If an experiment manipulated both physical cleansing and other psychological variables and measured outcomes directly about physical cleansing, then only effects pertaining to manipulations of other psychological variables were included.

It is worth noting that given our goal of providing a comprehensive meta-analysis of all cleansing effects, we allowed any form of physical cleansing to be included (e.g., thoughts about cleansing, desires for cleansing, scent of cleansing, evaluation of cleansing products, mental simulation of cleansing, actual behavior of cleansing).

(6) For each cleansing effect, we coded the original researchers' expectation and our expectation (i.e., the meta-analysts') about its significance. The original researchers either expected an effect to be (a) significant or (b) non-significant, or (c) had no clear expectation about it. Based on our own theoretical understanding, we (the meta-analysts) independently expected an effect to be either (a) significant, (b) non-significant, (c) significant but weaker than in other conditions of the same study, or (d) significant but in the opposite direction to the original researchers' expectation.

If an effect was expected by both the original researchers and the meta-analysts to be non-significant, it was excluded from the meta-analysis. If an effect was expected by either the original researchers or the meta-analysts or both to be significant, it was included. For example, an experiment examined the effect of physical cleansing on judgments of politicians (Kaspar & Klane, 2016). Some judgments were designed to be related to cleanliness; we included these in the meta-analysis. Other judgments were designed to be unrelated to cleanliness and were thus expected by both the original researchers and the meta-analysts to be non-significant; we excluded these from the meta-analysis.

In addition, there were experiments where the original researchers and meta-analysts held different expectations. For example, Briñol et al. (2013, Additional Experiment) expected that imagined movement of an electronic file to the recycle bin vs. the storage disk would be ineffective in shaping attitude (whereas they expected actual movement of the electronic file to the recycle bin vs. the storage disk to be effective in shaping attitude); we expected such imagined movement to be effective though weaker than actual movement. Pilotti and El Alaoui (2018, Experiment 1) expected that having participants recall unethical vs. ethical deeds would influence their choice between a cleansing and non-cleansing product 30 minutes later; we expected the lengthy delay between the manipulation and the measure to eliminate the effect. We included all of these instances in the meta-analysis.

Such inclusion criteria resulted in conservative estimates of overall effect sizes. There were six instances where the meta-analysts expected the effect to be significant but in the opposite direction to the original researchers' expectation. We coded the valence of these effect sizes on the basis of our own expectation. Four effect sizes ended up being negative, two positive.

(7) The report had to provide sufficient statistical information (e.g., means, *SD*s, cell sizes, t, Z, χ^2) for effect size computation. Where necessary, we contacted authors to obtain the required statistical information or estimated it from figures (e.g.,

measuring the length of an error bar). There were 31 reports for which we had to request necessary statistical information; we successfully obtained the necessary information for 22 of them. In addition to those 31 reports, there were two reports we requested but did not receive. If the necessary statistical information was available, the corresponding effect was included; otherwise, excluded.

(8) The report had to contain no ethical concern. Reports that had been retracted were excluded. To err on the conservative side, reports that involved authors with established records of data fraud were also excluded, even if there was no established evidence of fraudulent data with the report at hand.

Meta-Analytic Strategy

Computation of Effect Sizes

Effect sizes were computed using the Comprehensive Meta-Analysis (CMA) Version 2 software (Borenstein et al., 2005), which generated Cohen's *d* and Hedges' *g* as outputs. We used Hedges' *g* because it corrects for the tendency of Cohen's *d* to overestimate effect sizes in small samples (Borenstein et al., 2009; Hedges, 1981). To compute effect sizes, necessary statistical information—various combinations of cell sizes, means, *SD*s, sample size, difference in means, common *SD*, *SD* of mean difference, correlation between repeated measures, *t* test, Fisher's *Z*, χ^2, and *p* value—was extracted or estimated from original reports and entered as inputs to CMA. Effects were coded as positive (vs. negative) if they were consistent (vs. inconsistent) with our theoretical hypotheses (see ***Theoretical Types of Cleansing Effects***).

Effect sizes were computed and coded at the most specific level afforded by the reported data. If data were reported for multiple subgroups (e.g., when the manipulation of interest was crossed with another factor), we computed and coded the effect of interest for each subgroup. If data were reported for multiple measures (e.g., when both self-report and behavioral measures were taken), we computed and coded the effect for each measure of interest. We retained information about which effect belonged to which study in the meta-analytic dataset, which allowed us to conduct analyses at the effect-level (see ***Synthesis and Variability of Effect Sizes***) or by taking the nested nature of effects (k_{effect} = 551) within studies (k_{study} = 230) into account (see ***Handling of Non-Independent Effect Sizes***).

Synthesis and Variability of Effect Sizes

Using the *rma.uni* function in the *metafor* package 4.2.0 (Viechtbauer, 2010) in *R* 4.3.1 (R Core Team, 2023), we meta-analyzed the effect sizes in both a random-effects model and a fixed-effect model. In a random-effects model, effect sizes are weighted by $[1/(v_i + \tau^2)]$, where v_i is the within-effect error variance and τ^2 is the between-effects variance (i.e., residual heterogeneity) estimated with the restricted maximum-likelihood estimator (REML; Raudenbush, 2009; Viechtbauer, 2005). In a fixed-effect model, effect sizes are simply weighted by the inverse of their error variance (Borenstein et al., 2010).

In principle, we favored the use of random-effects over fixed-effect models in the present meta-analysis for several reasons (Borenstein et al., 2010; Hunter & Schmidt, 2000). (a) The meta-analyzed studies involved different manipulations, measures, and moderators. As such, we did not assume the studies to share a common effect size (which would be the assumption of a fixed-effect model). Instead, we assumed their effect sizes to be a random sample of an underlying distribution of true effects, and estimated the mean of these true effects, consistent with the assumptions of a random-effects model. (b) We were interested in generalizing the estimated effects beyond the experimental contexts observed thus far (consistent with the assumption of a random-effects model). (c) The number of studies was large enough. Had it been too small, the random-effects model would have been inappropriate.

In practice, we report meta-analytic results from both the random-effects model (in the paper) and the fixed-effect model (in the supplemental material). The reason is that we were concerned about publication bias, which tends to be manifest in the form of larger effect sizes in studies with smaller samples (Greenhouse & Iyengar, 2009). Mathematically, studies with small samples are assigned more weight in a random-effects model than in a fixed-effect model, whereas studies with large samples are assigned less weight in a random-effects model than in a fixed-effect model (Borenstein et al., 2010). As a corollary, if publication bias indeed existed in the form of larger effect sizes in studies with smaller samples, then the estimated mean of true effects in a random-effects model would be larger than the estimated common effect in a fixed-effect model. We decided to provide readers with both kinds of estimates, even though in principle only the random-effects model fits our theoretical assumptions and inferential goals.

To examine heterogeneity of effect sizes, for each model (whether random-effects or fixed-effect) we computed the homogeneity statistic Q with a χ^2 distribution and degrees of freedom $df = k - 1$, where k is the number of effect sizes at either the effect-level (k_{effect}) or study-level (k_{study}). A significant Q indicates significant heterogeneity of effect sizes (Borenstein et al., 2009). In addition, for each random-effects model, we estimated the amount of total heterogeneity between effects (τ^2) and computed the percentage of total variability due to heterogeneity between effects (I^2). Whereas the τ^2 statistic can be inaccurate when the number of effect sizes is small, the I^2 statistic is more consistent regardless of the number of effect sizes and is thus more comparable across meta-analyses (Higgins et al., 2003; Higgins & Thompson, 2002).

Handling of Non-Independent Effect Sizes

For a study with multiple effects, retaining all effect sizes attains comprehensiveness but violates the statistical assumption of independent effect sizes—an assumption underlying the traditional random-effects and fixed-effect models outlined above. To adjust for dependency among effect sizes within a study, we fit random-effects models with robust variance estimates (RVE; Hedges et al., 2010) in synthesizing effect sizes and conducting moderator analyses. We used the *R* package *robumeta* 2.1 (Fisher et al., 2023), with or without making small-sample corrections for the residuals and degrees of freedom (Tipton, 2015). We used the package's default value of rho (correlation among effect sizes within study) = .80 but also conducted sensitivity analysis to ascertain the robustness of results across different values of rho (0, .20, .40, .60, .80, and 1.00). Results included an intercept (precision-weighted overall effect size, adjusting for correlations among dependent effect sizes within study), its standard error, *t*-value, *p*-value, and 95% confidence interval.

The advantage of the RVE approach is that it does not require any knowledge about the correlation structure underlying the dependent effect sizes (whereas multivariate meta-regression does; Raudenbush et al., 1988). Nor does it entail any loss of information about different levels of a moderator within a study (whereas aggregating effect sizes within a study does; Borenstein et al., 2009; Rosenthal & Rubin, 1986). The disadvantage is that the RVE approach, together with high degrees of heterogeneity, limits the statistical power of testing for moderation, which often requires many observations to provide sufficient power (Hedges & Pigott, 2001, 2004). Therefore, null effects in moderator analyses using RVE should be treated with caution (Coles et al., 2019). For this reason, our narrative summary of the moderator analyses will focus on results from random-effects models without using RVE, though parallel results from random-effects models using RVE will be included in all relevant tables for transparency and ease of comparison.

If a moderator was continuous, it would be straightforward to enter it as a predictor in the meta-regression with RVE. But most of our moderators were categorical. If a categorical moderator had only two levels, it would be a dummy-coded predictor in the meta-regression with RVE, and its significance test would indicate whether the synthesized effect estimates

differed significantly between the two levels of the moderator. If a categorical moderator had more than two levels, it would be dummy-coded into more than one predictor in the meta-regression with RVE, where each predictor had its own significance test (for a single level against another single level). To obtain the omnibus significance test, we used the *Wald_test* function in the *R* package *clubSandwich* 0.5.8 (Pustejovsky, 2022) to conduct an approximate Hotelling-Zhang test with small-sample corrections (Tanner-Smith et al., 2016), resulting in an *F*-ratio that probed for any significant difference among all levels of the moderator.

Assessment of and Correction for Outliers and Publication Bias

To assess and correct for the impact of potential outliers, we used two approaches. First, we computed a variety of leave-one-out diagnostics using the *influence.rma.uni* function in the *R* package *metafor* 4.2.0 (Viechtbauer, 2010), excluded the diagnosed outliers and influential cases[6], and reran the primary analyses. Second, we conducted a sensitivity analysis by excluding effects in descending order of effect size and rerunning the primary analyses. We also added a supplemental analysis by excluding any effect that involved any of the present meta-analysts[7] (k_{effect} = 27, k_{study} = 14) and both rerunning the primary analyses and reapplying the two approaches (leave-one-out diagnostics and sensitivity analysis).

To assess and correct for publication bias, we used a variety of traditional and state-of-the-art methods, including (1) normal-quantile plot, (2) funnel plot with Egger regression test, (3) weighted least squares (WLS) meta-regression with no intercept and weighted average of adequately powered (WAAP) studies (together known as "WAAP-WLS"), (4) precision-effect test (PET) and precision-effect estimate with standard errors (PEESE; together known as PET-PEESE), (5) *p*-uniform and *p*-uniform*, and (6) parameter selection model (PSM). These methods operate under the assumption that effect sizes are independent of each other. In reality, as already noted, effect sizes are often dependent (e.g., because multiple effects come from the same experiment). To take the dependency of effect sizes into account, we also (7) ran PET-PEESE with robust variance estimates and (8) aggregated dependent effect sizes and submitted the aggregated estimates to the same publication bias analyses.

Why did we employ this large number of bias assessment and correction methods? Because recent systematic comparisons among a subset of them—namely, WAAP-WLS, PET, PEESE, PET-PEESE, *p*-uniform, PSM, and the traditional random-effects model—found that each method has its strengths and weaknesses, and that no single method consistently outperforms others under all research conditions such as assumed severity of publication bias, assumed extent of questionable research practices, and amount of heterogeneity (Carter et al., 2019).[8] Following these methodologists' recommendations, we used a two-step procedure.

First, we used their online app (http://shinyapps.org/apps/metaExplorer) to evaluate which methods were "expected a priori to perform reasonably well under research conditions that are most plausible for the meta-analysis at hand (*method performance check*)" (Carter et al., 2019, p. 137). Second, for transparency and to facilitate comprehensive evaluation by other researchers, we proceeded to "compute meta-analytic estimates using all the included methods and compare them in order to evaluate the variability (or robustness) of conclusions (*sensitivity analysis*). This evaluation should respect the results from the method performance check and weight the methods accordingly" (p. 137). A summary of our method performance check,

---

[6] We would have liked to identify outliers based on meta-regressions with RVE. Such methods, however, are not yet available (Coles et al., 2019).

[7] We thank the action editor for this suggestion.

[8] For additional discussion of the limitations of PET, PEESE, and PET-PEESE under research conditions such as small samples and large heterogeneity in effects between studies, see Stanley (2017). For caveats against the performance of *p*-uniform under research conditions such as large heterogeneity, publication bias, and *p*-hacking, see van Aert, Wicherts, and van Assen (2016). For cautions on PSM, see McShane, Böckenholt, and Hansen (2016).

followed by an outline of all methods and our implementation of them, is provided in Supplemental Material A.

Results of the method performance check suggest that if we have to pick a method that performs reasonably well under both H1 and H0 across the research conditions of our meta-analyzed experiments, it would be PET-PEESE. According to its originators, PET-PEESE is a conditional estimator such that if the effect size estimated by PET is *not* statistically significant with a one-tailed alpha of .05, then it is treated as the final estimate, but if the effect size estimated by PET is statistically significant with a one-tailed alpha of .05, then the effect size estimated by PEESE is used as the final estimate instead (Stanley & Doucouliagos, 2014). Beyond PET-PEESE, for comprehensiveness and following the meta-analytic recommendations by Carter et al. (2019), we still employed all eight methods of bias assessment and correction, presented their results, but gave more weight to PET-PEESE in our interpretation. In addition, to adjust for publication bias throughout our moderator analyses, we applied PET-PEESE with robust variance estimates to each moderator and, if the moderator was categorical, to each of its levels. We applied PET-PEESE (rather than all eight methods) to the moderator analyses because it would be unrealistic and overwhelming for the reader to glean the results of 8 methods × 23 categorical moderators × 2-4 levels per categorical moderator (see ***Examination of Moderators***). That said, all data and code are publicly available for interested readers to examine in detail.

Examination of Moderators

We conducted a series of single-moderator and double-moderator analyses. All moderators were coded by the first and second authors. As mentioned in footnote [5], ongoing supervision and iterative discussion guided the processes of moderator coding. Any disagreement was noted and resolved by discussion.

Moderator values were standardized for analysis (Cohen et al., 2013). Moderator analyses were conducted using meta-regressions with RVE (see ***Handling of Non-Independent Effect Sizes***) and PET-PEESE with RVE (see ***Assessment of and Correction for Outliers and Publication Bias***). Recall that RVE comes with a number of advantages but also the disadvantage that it limits the statistical power of testing for moderation, which means null effects in moderator analyses using RVE should be treated with caution (Coles et al., 2019).

If an experiment contained multiple includable effects (e.g., due to multiple subgroups, multiple measures), all corresponding effect sizes were computed (see **Inclusion Criteria**) and treated as effects nested within study (see ***Handling of Non-Independent Effect Sizes***). For each effect (whenever possible) or each study, we coded a variety of theoretical and exploratory moderators, as described below.

Theoretical Types of Cleansing Effects

Directionality. We coded each effect as demonstrating either (1) a psychological consequence or (2) a psychological antecedent of cleansing. For brevity, we refer to this distinction as the directionality of a cleansing effect.

Psychological consequences of cleansing involved manipulating physical cleansing and measuring other psychological variables. For example, an experiment manipulated physical cleansing by having participants either use a cleansing product (a bottle of liquid soap) or merely examine it after making a free choice between two similarly attractive options (music albums), then measured their evaluation of the chosen and rejected options (S. W. S. Lee & Schwarz, 2010b, Experiment 1).

Psychological antecedents of cleansing involved manipulating other psychological variables and measuring outcomes directly about physical cleansing. For example, an experiment manipulated the psychological variable of morality by having participants recall

either their ethical or unethical behavior, then measured their choice between a cleansing product (an antiseptic wipe) and a non-cleansing product (a pencil) as a free gift (Zhong & Liljenquist, 2006, Experiment 3).

Coding this moderator served two analytic purposes. First, it allowed us to compare the strength and heterogeneity of psychological consequences vs. antecedents of cleansing across all experiments (single-moderator analyses). Second, it allowed us to examine whether both directions of cleansing effects showed similar patterns of moderation by other moderators (double-moderator analyses).[9]

Theoretical Basis. Regardless of the directionality of a cleansing effect (i.e., whether it was demonstrating a psychological consequence or antecedent of cleansing), we coded it in terms of whether it was theoretically assumed to be (1) content-based or (2) procedure-based.

Content-based cleansing effects were those where the psychological variable of interest was assumed by existing theories to have shared content with physical cleansing. For example, morality and religiosity are generally assumed to have conceptual associations with physical cleanliness and purity (Lakoff & Johnson, 1999; S. W. S. Lee & Schwarz, 2016; Preston & Ritter, 2012; West & Zhong, 2015).

Procedure-based cleansing effects were those where the psychological variable of interest was assumed to be related to physical cleansing not because of shared content but because of procedural reasons. Examples of such procedural reasons, in conceptual terms, included the mental procedure of psychological separation that is grounded in the experience of physical cleansing (S. W. S. Lee & Schwarz, 2021), procedural demands of social presence, and procedural fluency. In empirical terms, psychological states such as dissonance (S. W. S. Lee & Schwarz, 2010b), luck (A. J. Xu et al., 2012), and endowment (Florack et al., 2014) do not have any content overlap with physical cleansing, but can still be psychologically separated from the present self via physical cleansing. Cues of being observed can elicit positive behaviors in general, one of which is handwashing in the bathroom (Pfattheicher et al., 2018). Processing fluency can elicit many positive evaluations, one of which is evaluation of water as clean and pure (Cho, 2019). Across these examples, the psychological variables do not have any obvious shared content with physical cleansing.

It is worth noting that to maximize the informativeness of this moderator, we coded whether a cleansing effect was content-based or procedure-based in a mutually exclusive, dichotomous manner. For a cleansing effect to be considered procedure-based, its psychological variable must have no shared content with physical cleansing. If there was any content overlap, the cleansing effect was considered content-based—even if procedures (e.g., separating past from present) might also be at work. (We cannot think of any feasible way, in principle or in practice, to guarantee that content-based effects were devoid of procedures.) That means procedure-based effects had to be driven by procedure alone, whereas content-based effects could be driven by content alone or by both content and procedure.

Among content-based cleansing effects, when an observed effect was consistent (vs. inconsistent) with the expected effect (be it assimilation or contrast, as explained in the next section), its effect size was coded as positive (vs. negative). Among procedure-based cleansing effects, when an observed effect was consistent (vs. inconsistent) with theoretical expectations (e.g., physical cleansing would result in psychological separation), its effect size was coded as positive (vs. negative).

Content-Based Assimilation vs. Contrast. Each content-based cleansing effect was coded as (1) assimilation or (2) contrast, based on the polarity of physical cleansing and the psychological variable of interest (Table 1). Assimilation effects were those where the clean (vs.

[9] For the purpose of running double-moderator analyses to test if directionality interacted with other categorical moderators, there had to be no empty cell in the cross-tabulation. If there was any empty cell in the cross-tabulation, the categorical moderator was not analyzed.

unclean) pole was expected to be empirically linked to the pole of the psychological variable that is positively associated with cleansing, like moral (e.g., behaving more prosocially in a clean-scented room than in an unscented room; Liljenquist et al., 2010, Experiment 2) or religious (e.g., using more cleansing-related words in a word completion task after unscrambling sentences containing religious concepts than after unscrambling sentences containing neutral concepts; Preston & Ritter, 2012, Experiment 1). Contrast effects were those where the clean (vs. unclean) pole was expected to be empirically linked to the pole of the psychological variable that is *not* positively associated with cleansing, like immoral (e.g., being more likely to choose a cleansing product after recalling one's unethical behavior than after recalling one's ethical behavior; Zhong & Liljenquist, 2006, Experiment 3) or hedonistic (e.g., favoring a more hedonic snack after wiping one's hands than after using a pencil; Martins et al., 2015, Experiment 2).

Assimilation and contrast effects can be different empirical manifestations of the same underlying conceptual hypothesis. For example, an experiment required female participants to listen to a scenario and imagine being the woman in it receiving either a non-consensual or consensual kiss, then measured their extent of feeling dirty as well as their urge to wash themselves and their likelihood of actually doing so during a break (Herba & Rachman, 2007). Participants who imagined receiving a non-consensual (vs. consensual) kiss were expected to report both a higher level of feeling dirty (assimilation effect of sexual immorality on dirtiness) and a stronger urge to wash themselves and a higher likelihood of doing so (contrast effect of sexual immorality on cleansing desire and behavior). It should be obvious that these expected assimilation and contrast effects were simply different empirical manifestations of the same underlying conceptual hypothesis that a non-consensual kiss, as an instantiation of sexual immorality, was linked to dirtiness.

Why do we bother to distinguish between assimilation and contrast if they can be simply different empirical manifestations of the same underlying conceptual relation? One reason is that even when both of them support the same underlying conceptual hypothesis, they may involve different mental processes, according to a long history of attitude and social cognition research on assimilative and contrastive effects (Bless & Schwarz, 2010; Herr et al., 1983; Mussweiler et al., 2004; Schwarz & Bless, 1992; Sherif & Hovland, 1961; Strack et al., 1993). Another reason is that teasing them apart may reveal useful empirical nuances, a theme we will explicate in the **Discussion**.

Theoretical Domains and Subdomains of Cleansing and Psychological Experience

Subdomain of Physical Cleansing. We coded the subdomain of physical cleansing as about (1) clean vs. dirty/neutral in general (e.g., wiping vs. not wiping one's hands; Zhong & Liljenquist, 2006, Experiment 4), which constitutes the vast majority of effects, (2) clean vs. dirty money in particular (e.g., counting clean vs. dirty money; Yang et al., 2013, Experiment 2), which carries different psychological meanings from those of cleanliness vs. dirtiness in general, or (3) discarding vs. keeping objects (e.g., throwing a piece of paper into the trash can vs. checking it for grammar or spelling errors; Briñol et al., 2013, Experiment 1). We distinguished between these subdomains of physical cleansing because they draw on different theoretical assumptions (for details, see Briñol et al., 2013; Jiang & Gao, 2015; C. Kim & Huh, 2019; Yang et al., 2013). Empirically, subdomain 2 (clean vs. dirty money) was observed only when physical cleansing was manipulated (as the independent variable in psychological consequences of cleansing), whereas subdomains 1 and 3 were observed when physical cleansing was either manipulated (as the independent variable in psychological consequences of cleansing) or measured (as the dependent variable in psychological antecedents of cleansing).

Domain and Subdomain of Psychological Variable. To test if cleansing effects differed in strength within vs. beyond the moral domain, we coded the psychological variable in each effect as (1) directly related, (2) indirectly related, or (3) unrelated to morality.

If the cleansing effect involved a psychological variable that was (1) directly related to morality, we coded which moral subdomain it invoked (Graham et al., 2013): (a) care/harm (e.g., playing a video game with violence against humans; Gollwitzer & Melzer, 2012); (b) fairness/cheating (e.g., returning money in a one-shot anonymous trust game; Liljenquist et al., 2010, Experiment 1); (c) sanctity/degradation, the content of which was further coded as sexual (e.g., imagining borrowing a phone from a gay vs. heterosexual man; Golec de Zavala et al., 2014, Experiment 1), non-sexual (e.g., judging the wrongness of eating a dead dog; Schnall et al., 2008, Experiment 1), or both; or (d) other subdomains, mixed subdomains, or morality in general ("other/mixed/general").

The last category was a mixed bag because of the relatively small effect counts in other subdomains, namely, loyalty/betrayal (e.g., smelling a shirt belonging to an ingroup vs. outgroup member; Reicher et al., 2016, Experiment 2) and authority/subversion (e.g., receiving bonus based on subordinates' performance; Cramwinckel, De Cremer, et al., 2013, Main Experiment). Mixed subdomains or morality in general included:

- honesty/dishonesty (e.g., telling the truth vs. lying; S. W. S. Lee & Schwarz, 2010a);
- deontological vs. altruistic guilt (e.g., listening to a story about violating one's own moral rules vs. failing to help a victim; D'Olimpio & Mancini, 2014, Experiment 2);
- prosocial intention or behavior (e.g., being willing to help the experimenter; Kalanthroff et al., 2015);
- authenticity/inauthenticity (e.g., being willing to purchase a counterfeit t-shirt; J. Kim et al., 2018, Experiment 5); or
- when the experimental design did not focus on any particular subdomain (e.g., rating one's likelihood of committing morally bad actions in the future; Kaspar & Teschlade, 2016, Experiment 1).

If the cleansing effect involved a psychological variable that was (2) indirectly related to morality or (3) unrelated to morality, we further coded it as (a) social (i.e., involving others or society) or (b) non-social (i.e., not involving others or society).[10]

Social subdomains that were indirectly related to morality included:

- politics (e.g., rating one's political attitudes; Helzer & Pizarro, 2011, Experiment 1);
- religion (e.g., unscrambling sentences containing religion-related vs. neutral words; Preston & Ritter, 2012, Experiment 1);
- ostracism (e.g., receiving two vs. ten out of thirty ball tosses in Cyberball game; Poon, 2019, Experiment 1);
- attitude (prejudice towards outgroup; e.g., smelling a shirt belonging to an ingroup vs. outgroup member; Reicher et al., 2016, Experiment 2);
- empathy and self–other focus (e.g., seeing signs about patient vs. personal consequences of hand-hygiene; Grant & Hofmann, 2011, Experiment 1); and
- trust/suspicion (e.g., rating one's trust in a person described in a scenario; Basáñez et al., 2019, Experiment 2).

Social subdomains that were unrelated to morality included:

[10] Some effects involved both the moral domain and another domain. For example, an experiment had participants recall their unethical deed, use an antiseptic wipe (vs. not), and rate both moral and non-moral emotions (Zhong & Liljenquist, 2006, Experiment 4). The effect of cleansing after unethical recall on non-moral emotions involved both morality (in the manipulation) and emotion (in the measure). Accordingly, its manipulation was coded for which moral subdomain it involved and its measure was coded for whether it was social or non-social.

- emotion (e.g., rating one's amusement, calmness, confidence, etc.; Zhong & Liljenquist, 2006, Experiment 4);
- gender (e.g., being exposed to a picture of male vs. female eyes; King et al., 2016);
- social presence (e.g., being aware of the presence of others in the bathroom; Cardinale Lagomarsino et al., 2017);
- cooperation (e.g., amount of money offered in cooperation scenario; Schwader, 2013, Experiment 4);
- identity threat (e.g., reading an article that threatened the genetic distinctiveness of one's cultural community; Nussinson et al., 2019, Experiment 1); and
- mortality (e.g., completing questions that required detailed descriptions of death; Strachan et al., 2007, Experiment 2).

Non-social subdomains that were unrelated to morality included:

- postdecisional dissonance (e.g., making a free choice between two similarly attractive fruit jams; S. W. S. Lee & Schwarz, 2010b, Experiment 2);
- luck (e.g., having good vs. bad luck in gambling; A. J. Xu et al., 2012, Experiment 2);
- optimism (e.g., rating optimism about one's future performance on an intellectual task; Kaspar, 2012);
- endowment (e.g., being given a specific soft drink; Florack et al., 2014, Experiment 1);
- attitude (towards object; e.g., writing down positive vs. negative thoughts about the Mediterranean diet; Briñol et al., 2013, Experiment 2);
- attitude (towards self; e.g., writing down positive vs. negative thoughts about one's own body; Briñol et al., 2013, Experiment 1);
- information processing (e.g., reaction time in the Stroop task; Kalanthroff et al., 2015);
- stress (e.g., rating one's perceived stress within the last month; Kaspar & Cames, 2016, Experiment 1);
- fluency (e.g., easy- vs. hard-to-pronounce brand names; Cho, 2019);
- goal activation (e.g., unscrambling sentences containing words about the importance of academic achievement; Ma & Lee, 2022, Experiment 1);
- change (e.g., writing about ways in which one wanted to change one's life; Jiang & Gao, 2015, Experiment 1b);
- ownership (e.g., using Lego bricks to build a robot oneself; A. Lee & Ji, 2015);
- risk taking (e.g., imagining volunteering at a hospital ward with pathogens and infection present; Prokosch et al., 2019, Experiment 2);
- saving/spending (e.g., allocating hypothetical bonus from work to mortgage payment versus spa vacation; Morrin et al., 2014);
- shape (e.g., seeing an ad for product collection groups in circle vs. square; Jia et al., 2018, Experiment 3); and
- threat specificity (e.g., seeing an ad on washing hands to prevent sickness in general vs. COVID-19 in particular; Sobol & Giroux, 2021, Experiment 3).

One non-social subdomain that we considered indirectly related to morality was healthy/unhealthy eating (e.g., number of unhealthy food choices; J. Kim et al., 2018, Experiment 1) because it is often construed as an issue of self-control (Fitouchi et al., 2022), and self-control is moralized (Hofmann et al., 2018; Mooijman et al., 2018; Weiss et al., 2021).

Operationalizations of Manipulation and Measure

Operationalization of Manipulation. Regardless of whether an effect involved manipulating physical cleansing (in psychological consequences of cleansing) or other

psychological variables (in psychological antecedents of cleansing), we coded the manipulation as (1) actual experience, (2) imagined or recalled experience (i.e., mentally simulated experience), or (3) merely conceptual activation (i.e., conceptual activation without involving any actual experience or imagined or recalled experience). For example, physical cleansing could be manipulated by having participants use vs. not use an antiseptic wipe (actual experience; Zhong et al., 2010, Experiment 1), read and visualize a statement written in first person about being clean vs. dirty (mentally simulated experience; Experiment 2), or unscramble sentences containing cleansing-related vs. neutral words (merely conceptual activation; Schnall et al., 2008, Experiment 1). Other psychological variables could be manipulated by having participants tell a truth vs. lie on voicemail (actual experience; S. W. S. Lee & Schwarz, 2010a), copy a story written in first person about helping vs. sabotaging a coworker (mentally simulated experience; Zhong & Liljenquist, 2006, Experiment 2), or unscramble sentences containing religion-related vs. neutral words (merely conceptual activation; Preston & Ritter, 2012, Experiment 1).

Operationalization of Measure. Regardless of whether an effect involved measuring other psychological variables (in psychological consequences of cleansing) or physical cleansing (in psychological antecedents of cleansing), we coded the measure as involving (1) behavior, (2) judgment or feeling, or (3) thought or sensorimotor process.

Measures of behavior included:

- social or economic behavior (e.g., amount of money donated to charity; Lobel et al., 2015, Experiment 2);
- nonsocial, noneconomic choice behavior (e.g., choice between keeping an endowed soft drink and exchanging it for another soft drink; Florack et al., 2014, Experiment 1); and
- physical behavior (e.g., amount of hand gel used from dispensers; Grant & Hofmann, 2011, Experiment 1).

Measures of judgment or feeling included:

- judgment or feeling about oneself (e.g., self-rated guilt and shame; H. Xu et al., 2014);
- judgment or feeling about others (e.g., dirtiness of a transgressor; Rothschild et al., 2015, Experiment 2); and
- judgment or feeling about objects or the environment (e.g., desirability of toothpaste and mouthwash products; Schaefer et al., 2015).

Measures of thought or sensorimotor process included:

- concept accessibility (e.g., word completion tasks using cleansing-related words; Zhong & Liljenquist, 2006, Experiment 1);
- cognitive processes, which ranged from memory (e.g., recognition memory of moral social issues; Kaspar & Jahn, 2016) and intellectual performance (e.g., number of anagrams solved; Kaspar, 2012) to task completion (e.g., response time in evaluating mouth-cleansing products; Denke et al., 2014); and
- sensorimotor processes with clear psychological referents in the context of the experiment at hand, such as response time in stop-signal trials (Kalanthroff et al., 2015), change in pupil size when viewing immoral scenes (Kaspar et al., 2015), and change in heart rate variability (Ottaviani et al., 2018). In general, however, neural and physiological measures have multiple mappings with psychological constructs (Cacioppo et al., 2007) so they are included in analysis only if they have clear psychological referents in the context of the experiment at hand.

Experimental and Statistical Design

Was the Cleansing Effect Tested for Mediation by Another Measure? For each cleansing effect, we coded whether it (1) was or (2) was not tested for mediation by another measure and, if so, whether mediation (1) was or (2) was not supported by the evidence. As an

example of mediation being tested and supported, participants who were reminded (vs. not reminded) to use an antiseptic wipe judged moral violations of sexual purity norms more harshly, an effect that was mediated by more politically conservative attitudes (Helzer & Pizarro, 2011, Experiment 2; but see Burnham, 2020). As an example of mediation being tested and not supported, participants who visualized being in a clean (vs. dirty) bodily state perceived lower levels of stress, an effect that was tested for mediation by how long participants took in a prior stress recall task. The mediating role was not supported (Kaspar & Cames, 2016, Experiment 2).

Did the Study Test Content-Based Specificity Using Another Measure? For each study that was theoretically assumed to be content-based, we coded whether content-based specificity (1) was or (2) was not tested and, if so, whether it (1) was or (2) was not supported by the evidence. For content-based specificity to be tested, a study had to include not only the focal measure that was expected to show the cleansing effect, but also at least one additional measure that was expected to *not* show the same effect (i.e., a significant effect in a different direction or a non-significant effect). For content-based specificity to be supported, the focal measure had to show the expected cleansing effect, whereas the additional measure had to *not* show the same effect. The additional measure could be in the same or a different content domain.

As an example of content-based specificity being tested and supported by an additional measure in a different domain, asking (vs. not asking) participants to use an antiseptic wipe after they recalled their own unethical behavior reduced their negative moral emotions, but did not influence their non-moral emotions (Zhong & Liljenquist, 2006, Experiment 4). The focal measure (negative moral emotions) was within the domain of morality; the additional measure (non-moral emotions) was not. As an example of content-based specificity being tested and supported by an additional measure in the same domain, asking (vs. not asking) participants to use an antiseptic wipe increased recognition memory of immoral issues, but decreased recognition memory of moral issues (Kaspar & Jahn, 2016). Both the focal measure (immoral issues) and the additional measure (moral issues) were within the domain of morality.

For studies that were theoretically assumed to be procedure-based, content-based specificity was irrelevant and thus not coded.

Did the Study Test Moderation by Another Factor? Beyond content-based specificity, for each study we coded whether the cleansing effect (1) was or (2) was not tested for moderation by at least one additional factor and, if so, whether moderation (1) was or (2) was not supported by the evidence. As an example of moderation being tested and supported, participants who were reminded of death (vs. dental pain) subsequently spent more time washing their hands and used more soap—an effect that was found only among obsessive-compulsive disorder (OCD) washers, not among OCD non-washers (Menzies & Dar-Nimrod, 2017, Experiment 2). As an example of moderation being tested and not supported, participants who washed their hands (vs. rubbed their hands with chocolate) showed less postdecisional change in ranked preference for a chosen option over a rejected option (Marotta & Bohner, 2013). This effect was tested for moderation by individual differences in preference for consistency. The moderating role was not supported.

Did the Focal Factor Include Another Condition? For each cleansing effect, we coded whether the focal factor (1) did or (2) did not involve any additional condition beyond the two focal conditions. If the focal factor involved any additional condition, we only included data from the two conditions expected to show the strongest cleansing effect and did not include data from any additional condition. This analytic strategy guarded us against "double-dipping" the data (i.e., using the same data point in the meta-analysis more than once; Senn, 2009). For example, an experiment assigned participants to one of three conditions (Yang et al., 2013, Experiment 6): reading a news article about how clean the country's paper currency was, or how filthy it was, or reading a recent weather report. We only included data from the two focal conditions (clean and filthy), but not data from the additional condition (weather report). Had we included data from the additional condition, we would have had to compare it with either or

both of the focal conditions, thereby double-dipping the data. We avoided this problem.[11] The goal of coding this moderator (presence of additional conditions) was simply to inform readers of studies that contained additional data for their own exploratory interest.

Was the Study Preregistered? For each study, we coded whether its data collection and analytic plan (1) had or (2) had not been preregistered.

Participant Demographics

Type of Participants. For each study, we coded whether it involved (1) local or university-related participants (students, university community, local community), (2) online participants, or (3) unspecified or other types of participants (professional association, company employees, factory workers, political activists, patients with obsessive-compulsive disorder, healthcare practitioners, hospital community).

Region of Participants. For each study, we coded its participants' geographical region: (1) West, including Australia, Austria, Canada, France, Germany, Italy, Luxembourg, Netherlands, Poland, Portugal, Spain, the U.K., the U.S., and North America in general; (2) East, including China, Hong Kong, India, Japan, South Korea, and Taiwan; (3) Middle East, including Israel, Saudi Arabia, Turkey, and United Arab Emirates; or (4) Other, including South America (Argentina), a mix of East and West, global, and unspecified.

Gender Ratio of Participants. For each study, we coded the female percentage of its participants, ranging from 0% to 100%. Some studies did not report any gender information.

Report Characteristics

Year of Report. For each report, we coded its year of publication or availability.[4] It ranged from 1989 to 2023.

Type of Report. For each report, we coded whether it (1) was or (2) was not peer-reviewed. The former included peer-reviewed journal articles and conference presentations. The latter included unpublished reports, theses, and internal conference presentations.

How Report Was Presented by Authors. We coded whether each report was presented by its authors as (1) an original experiment, (2) a successful replication, or (3) an unsuccessful replication.

Transparency and Openness

To enhance meta-analytic transparency and reproducibility (Polanin et al., 2020), all of our data, code, and research materials (including our coding scheme) have been made publicly available at the Open Science Framework and can be accessed at https://osf.io/xv2wj/. Data were analyzed using *Comprehensive Meta-Analysis* Version 2 (Borenstein et al., 2005) and *R* 4.3.1 (R Core Team, 2023) with the aid of various packages noted throughout our paper. This review was not preregistered, but we have strived to satisfy all other relevant criteria on the AMSTAR 2 inventory[12] (Shea et al., 2017). We also adhere to all PRISMA (Moher et al., 2009) and MARS (American Psychological Association, 2020; Appelbaum et al., 2018, pp. 21–23)

---

[11] In this particular example (Yang et al., 2013, Experiment 6), had we included all three conditions, the "clean vs. filthy" comparison and the "clean vs. control" comparison would have been most relevant to our theoretical focus. The "filthy vs. control" comparison would not have been as relevant because it would be testing the effect of filthy, not the effect of clean (see **Overview of the Present Meta-Analysis**).

[12] We should point out that the AMSTAR 2 inventory is intended as "a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of healthcare interventions" (Shea et al., 2017, p. 1). Although our meta-analysis is not about healthcare interventions, nor does it include non-randomized studies, we do find the AMSTAR 2 inventory useful as a critical appraisal tool for systematic reviews of psychological research (cf. B. T. Johnson, 2021; B. T. Johnson & Hennessy, 2019).

guidelines for meta-analytic reporting. Our self-assessments using both the AMSTAR 2 inventory and the PRISMA checklist are provided in the supplemental material.

Results

Description of the Studies

As noted in **Method**, we coded a variety of moderators, from theoretical and methodological ones to participant demographics and report characteristics. Almost all of the moderators were categorical. Effect count at each level of each categorical moderator is presented in Table S1 (first column of results, total k_{effect} = 551). Descriptions of the effect counts at different levels of the theoretical and methodological moderators will be provided in conjuction with their effect sizes in the **Moderator Analyses** to optimize contextual understanding of which theoretical and methodological kinds of cleansing effects have received the most attention and shown the strongest effects. The present section provides a brief summary of effect counts based on moderators pertaining to participant demographics and report characteristics in hopes of sketching the empirical landscape covered by the meta-analytic data.

Most effects were based on local or university-related samples (k_{effect} = 396), followed by online (k_{effect} = 107) and other or unspecified samples (k_{effect} = 48). Most effects were based in the West (k_{effect} = 423); some were based in the East (k_{effect} = 71), Middle East (k_{effect} = 31), and other or unspecified regions (k_{effect} = 26). Most effects were reported in peer-reviewed journal articles (k_{effect} = 425); some were reported in peer-reviewed conference presentations (k_{effect} = 45), theses (k_{effect} = 38), unpublished reports (k_{effect} = 35), and internal conference presentations (k_{effect} = 8). The vast majority of effects came from reports presented by their authors as original experiments (k_{effect} = 470); far fewer effects came from reports presented by their authors as successful replications (k_{effect} = 32) or unsuccessful replications (k_{effect} = 49).

In addition to these categorical moderators, two of the moderators pertaining to participant demographics and report characteristics were continuous: female percentage of participants (study-level M = 60.88%, SD = 19.95%; Min = 0%, Mdn = 59.2%, Max = 100%, Range = 100%) and year of report (report-level M = 2014.35, SD = 4.55; Min = 1989, Mdn = 2014, Max = 2023, Range = 34). Their distributions are depicted in Figures S1 (female percentage of participants) and S2 (year of report).

Synthesized Effect Size

Table 2 summarizes the synthesized effect size estimate (g) and related statistics based on the three meta-analytic models: random-effects (RE) model, RE model using robust variance estimates (RVE) with small-sample corrections (SSC), and RE model using RVE without SSC. (Supplemental results based on the fixed-effect [FE] model are presented in Table S2.) Across the RE models, the synthesized effect size estimate was in the small-to-medium range (g = 0.411 to 0.455, ps ≤ 1.22E-40). Considerable heterogeneity was observed (τ^2 = 0.198, SE = 0.016; Q = 2669.709, p = 1.53E-274), as is typical of meta-analyses in general and expected from the variety of theoretical foci, domains, subdomains, manipulations, measures, and samples involved in our meta-analysis in particular. Heterogeneity between effects accounted for 84.767% of the total heterogeneity (I^2), indicating a large amount of non-random variability (Borenstein et al., 2009; Higgins et al., 2003; Higgins & Thompson, 2002; Huedo-Medina et al., 2006; Viechtbauer, 2010) that merits moderator analyses. The distribution of effect sizes is visualized in a funnel plot (Figure S3a), a histogram (Figure S3b), and a stem-and-leaf plot (Figure S3c), all of which suggest the need to address outliers[13] and publication bias.

[13] For example, the largest effect size (g = 11.565) came from a study where the manipulation involved female participants listening to a scenario and imagining being the woman in it, receiving a kiss

Outliers and Publication Bias

Recall that we addressed outliers in two ways. (1) We diagnosed outliers and influential cases based on the RE model (Figure S4a) or the FE model (Figure S4b) using the *influence.rma.uni* function in the *R* package *metafor* 4.2.0 (Viechtbauer, 2010). We excluded the identified outliers and influential cases, then reran the primary analyses. Results based on the RE models are summarized in Table 2 and visualized in Figures S5a-S5b and S6a-S6b. (Supplemental results based on the FE model are summarized in Table S2 and visualized in Figures S5c-S5d and S6c.) Across the RE models, the synthesized effect size estimate remained in the small-to-medium range (g = 0.385 to 0.393, ps ≤ 2.67E-42). But the diagnostics failed to exclude some obvious outliers, calling for another way to address them.

(2) We thus conducted a sensitivity analysis by excluding effects in descending order of effect size (excluding gs > 4.0, 2.0, 1.5, 1.0, or 0.5) and rerunning the primary analyses. Results are summarized in Table 2 and visualized in Figures S5e-S5n. (Supplemental results based on the FE model are summarized in Table S2.) With the most dramatic exclusion (excluding gs > 0.5), the total number of effects would drop by 42.3% from 551 to 318, the distribution of effect sizes would be truncated with a left skew (Figure S5n), and the synthesized effect size estimate would be in the small range and remain statistically significant regardless of which specific RE model was used (g = 0.143 to 0.160, ps ≤ 6.33E-15).

We also added a supplemental analysis by first excluding any effect that involved any of the present meta-analysts (k_{effect} = 27, k_{study} = 14), then both rerunning the primary analyses (last section) and reapplying the two approaches to addressing outliers (this section). All conclusions remain the same (Tables S3-S4).

These results suggest that the overall presence of cleansing effects was unlikely to be a mere consequence of outliers. It might be due to publication bias though, which we assessed and corrected for using an array of traditional and contemporary methods. We applied these methods either without excluding any outliers or after excluding outliers. Because diagnostics was insufficient for excluding some obvious outliers, we also examined the publication bias results using a sensitivity-analysis approach, excluding effects in descending order of effect size (excluding gs > 4.0, 2.0, 1.5, 1.0, or 0.5). Key results are presented in Table 3. Full results are available in Table S5.

Normal-quantile plots (Figures S6d-S6h) and Shapiro-Wilk normality tests (Table 3, Method 1) indicated that the distribution of effect sizes deviated positively from normality if gs > 4.0 or 2.0 were excluded, but did not deviate significantly from normality if gs > 1.5 were excluded, and deviated negatively from normality if gs > 1.0 or 0.5 were excluded. This pattern suggested that excluding gs > 1.5 might be a reasonable criterion, but results from additional publication bias analyses suggested that excluding gs > 1.0 or 0.5 might be a better criterion. Specifically, Egger regression tests (Table 3, Method 2) found a significant slope if gs > 4.0, 2.0, 1.5, or 1.0 were excluded, but a non-significant slope if gs > 0.5 were excluded, thus favoring the criterion of excluding gs > 0.5. With this strictest exclusion criterion, however, the distribution

---

from her boyfriend who later betrayed her vs. receiving a kiss from her boyfriend and later witnessing a kiss between two strangers. The measure was the change in dirty feelings from pre- to post-manipulation.

Three of the next largest effect sizes came from a single study. Two of these effect sizes involved participants with obsessive-compulsive disorder (OCD) writing about their unethical behavior, then wiping vs. not wiping their hands, and indicating their willingness to volunteer help (g = 6.816) and rating their moral emotions (g = 5.296). An additional effect size involved participants without OCD writing about their unethical behavior, then wiping vs. not wiping their hands, and rating their moral emotions (g = 4.312). In our process of reaching out to the authors to check if these effect sizes were accurate, we were told that the effect sizes reported in their paper contained clerical errors and that the effect sizes we computed are correct.

would show a clear truncation with a left skew (Figures S5m-S5n). With the next strictest exclusion criterion (excluding *g*s > 1.0), there would be no clear truncation (Figures S5k-S5l).

Balancing these considerations, we proceeded by applying this criterion (excluding effects with $g > 1.0$, or equivalently, including effects with $g \leq 1.0$, which also meant retaining null effects with $g \sim 0$ and effects contrary to hypothesis with $g < 0$) to our subsequent analyses ($k_{effect} = 479$). It would have been desirable to have our approach to deciding the exclusion criteria preregistered. In hopes of mitigating concerns about the robustness of our results, we maximize transparency in reporting both effect counts and effect sizes: Figure S7 is a stacked bar chart that shows the count of included effects (with $g \leq 1.0$) and excluded effects (with $g > 1.0$) at each level of each categorical moderator so that interested readers can easily examine the theoretical and methodological properties of each set of effects.[14] Table 3 reports the key results (and Table S5 reports the full results) of publication bias analyses based on all effects ("No exclusion" column) as well as based on effects after applying each of the possible exclusion criteria (other columns). A narrative summary of the results of all publication bias analyses is provided in Supplemental Material B.

All together, various methods of assessing and correcting for publication bias suggested that even after accounting for outliers (excluding effects with $g > 1.0$) and publication bias, the synthesized effect size estimate was generally significant and in the small range ($g = 0.103$ to 0.331). In other words, outliers and publication bias were unlikely to fully account for the overall presence of cleansing effects. At the same time, all methods found high degrees of heterogeneity between effects, suggesting the need to probe for moderators, which will provide more informative insights than the overall publication bias tests can afford (B. T. Johnson, 2021; B. T. Johnson & Hennessy, 2019).

Moderator Analyses

Table 4 shows the omnibus significance test for each categorical or continuous moderator. Figure 2 depicts the estimated effect sizes at different levels of each categorical moderator based on the RE meta-regression using RVE with SSC. Further results based on PET and PEESE (to correct for publication bias within moderator analyses) using RVE with SSC are also available in Table 4 and depicted in Figures S8-S9. For each categorical moderator, the funnel plot is available in Figure S10, and the statistical details for both the simple effects of all levels (comparison against $g = 0$) and the pairwise comparisons between all levels are available in the supplemental material on the Open Science Framework.

Based on the RE meta-regression using RVE with SSC, cleansing effects were similar in size across many moderators, but differed significantly in size between different subdomains of physical cleansing; between different operationalizations of measure; between studies that did vs. did not support content-based specificity; between studies that were vs. were not

[14] One observation about the excluded effects is particularly noteworthy. Various experiments asked female participants to imagine unwelcomed sexual encounters (e.g., non-consensual kiss) and measured their cleansing-related thoughts, desires, and behaviors. These experiments often found large effect sizes with $g > 1.0$ and ended up being excluded from our meta-analytic dataset. The large effect sizes might be expected though on the basis of content overlap between sexual morality and cleanliness/purity (Graham et al., 2013). Against this theoretical backdrop, it might be debateable whether these effects should be excluded from our meta-analytic dataset just because of their large effect sizes. But one could also argue that for any large effect sizes, compelling post hoc stories might be told to justify them if the meta-analyst is creative enough. Coupled with the publication bias analyses we conducted, we decided to err on the conservative side by excluding large effect sizes regardless of our own judgment of how plausible it was to provide theoretical justification for them. The implications of our conservative decision should be taken into account by researchers who wish to use the present meta-analytic findings to inform their future study design. For example, if they want to pick a psychological manipulation with large effects on cleansing desires, sexual violations might be worth considering.

preregistered; between different regions of participants; between different types of report; between reports that were presented by their authors as original experiments, successful replications, or unsuccessful replications; and as a function of the year of the report. A brief and systematic summary of the statistical results is presented below in a purely descriptive fashion, saving all interpretation (i.e., the "what does it all mean?") for the **Discussion**.

Which Theoretical Types of Cleansing Effects Showed the Strongest Effects?

Directionality. Effect sizes were statistically comparable between the psychological consequences (g = 0.320, p = 2.73E-21) and antecedents (g = 0.310, p = 4.21E-17) of cleansing, $F(1, 175.591) = 0.074$, p = .786. Effect counts, however, were considerably higher for the psychological consequences (k_{effect} = 296) than antecedents (k_{effect} = 183) of cleansing, $\chi^2(1) = 26.658$, p = 2.43E-07.

Theoretical Basis. Effect sizes were comparable between content-based (g = 0.329, p = 7.12E-32) and procedure-based (g = 0.274, p = 3.39E-07) cleansing effects, $F(1, 64.237) = 1.251$, p = .268. There were many more content-based (k_{effect} = 399) than procedure-based (k_{effect} = 80) cleansing effects, $\chi^2(1) = 212.445$, p = 4.02E-48. These results support both content and procedural aspects of cleansing, even though far more empirical attention has been paid to the former than the latter.

Content-Based Assimilation vs. Contrast. Effect sizes were comparable between content-based assimilation (g = 0.344, p = 4.09E-18) and contrast (g = 0.315, p = 1.17E-16) effects, $F(1, 149.409) = 0.496$, p = .482. There were similar numbers of content-based assimilation (k_{effect} = 205) and contrast (k_{effect} = 196) effects, $\chi^2(1) = 0.202$, p = .653. In other words, both assimilation and contrast are well supported by experimental work on cleansing effects.

What Domains of Psychological Experience and Subdomains of Cleansing Showed the Strongest Cleansing Effects?

Subdomain of Physical Cleansing. Effect sizes differed significantly between different subdomains of physical cleansing, $F(2, 7.398) = 15.346$, p = .002. Specifically, effect sizes were larger for the subdomain "clean vs. dirty money" (g = 0.731, p = .000194) than for "discarding vs. keeping objects" (g = 0.399, p = .00686E-3; $F(1, 9.046) = 7.443$, p = .0232) or "clean vs. dirty/neural in general" (g = 0.302, p = 9.55E-34; $F(1, 4.909) = 34.448$, p = .00217). Effect sizes did not differ significantly between the last two subdomains, $F(1, 6.538) = 0.907$, p = .375.

Despite the larger effect sizes for "clean vs. dirty money" than for the other subdomains of physical cleansing, we hasten to note that all of the effects for "clean vs. dirty money" (k_{effect} = 10) came from a single report (Yang et al., 2013) and that all of the effects for "discarding vs. keeping objects" (k_{effect} = 9) came from three reports (Briñol et al., 2013; Jiang & Gao, 2015; C. Kim & Huh, 2019). Both of these subdomains were vastly outnumbered by the broad category of "clean vs. dirty/neutral in general" (k_{effect} = 460). Even though it is useful to distinguish between these subdomains of physical cleansing (because they draw on different theoretical assumptions; Briñol et al., 2013; Jiang & Gao, 2015; C. Kim & Huh, 2019; Yang et al., 2013), the lopsided distribution of their effect counts ($\chi^2(2) = 847.395$, p = 9.79-185) suggests that caution is warranted in interpreting the difference in their effect sizes.

Domain and Subdomain of Psychological Variable. Effect sizes were comparable regardless of whether the psychological domain was directly related to morality (g = 0.317, p = 1.62E-23), indirectly related to morality (g = 0.387, p = 1.11E-09), or unrelated to morality (g = 0.273, p = 4.02E-08), $F(2, 77.368) = 1.761$, p = .179. These results suggest that cleansing effects are observable across domains, even though empirical work has been unevenly distributed between these domains ($\chi^2(2) = 269.390$, p = 3.18E-59), with a much stronger focus within

morality (326 effects directly related to morality, 49 effects indirectly related to morality) than beyond (104 effects unrelated to morality).

Effects that were directly related to morality tapped into different moral subdomains, including care/harm (k_{effect} = 21), fairness/cheating (k_{effect} = 60), sanctity/degradation (k_{effect} = 92), and other or mixed subdomains or morality in general (k_{effect} = 153). Effect sizes were comparable between them (care/harm g = 0.186, p = .0341; fairness/cheating g = 0.288, p = .000288; sanctity/degradation g = 0.291, p = 2.85E-07; "other/mixed/general" g = 0.349, p = 2.22E-16), $F(3, 35.689) = 1.486$, p = .235. Within the moral subdomain of sanctity/degradation, there were many more effects involving sexual content (k_{effect} = 71) than effects involving non-sexual content (k_{effect} = 11) or effects involving both sexual and non-sexual content (k_{effect} = 10), $\chi^2(2) = 79.587$, p = 5.22E-18, but effect sizes were comparable between them (sexual g = 0.280, p = 1.78E-05; non-sexual g = 0.225, p = .00796; both g = 0.273, p = .0389), $F(2, 10.205) = 0.263$, p = .774.

Effects that were indirectly related or unrelated to morality tapped into social (k_{effect} = 90) and non-social subdomains (k_{effect} = 94), with comparable effect counts ($\chi^2(1) = 0.0870$, p = .768) and effect sizes (social g = 0.336, p = 1.77E-11; non-social g = 0.314, p = 2.07E-08; $F(1, 84.174) = 0.135$, p = .714). These results consistently reflect the domain- and subdomain-general nature of cleansing effects.

What Operationalizations of Manipulation and Measure Showed the Strongest Cleansing Effects?

Operationalization of Manipulation. There were many more effects where the manipulation involved actual experience (k_{effect} = 219) or imagined or recalled experience (k_{effect} = 201) than conceptual activation (k_{effect} = 59), $\chi^2(2) = 96.217$, p = 1.28E-21. But effect sizes were comparable between them, $F(2, 72.489) = 0.735$, p = .483, suggesting that cleansing effects can be obtained by on-line experience (g = 0.344, p = 8.71E-19), mentally simulated off-line experience (g = 0.293, p = 2.00E-15), or merely conceptual activation (g = 0.309, p = 5.76E-06).

Operationalization of Measure. Effect sizes differed significantly between different operationalizations of measure, $F(2, 69.252) = 5.495$, p = .006. Specifically, effect sizes were larger for measures involving behavior (g = 0.409, p = 7.37E-19) than for measures involving judgment or feeling (g = 0.276, p = 8.42E-19; $F(1, 121.867) = 10.428$, p = .00160) and for measures involving thought or sensorimotor process (g = 0.269, p = .000149; $F(1, 44.160) = 4.194$, p = .0465). Effect sizes did not differ significantly between measures involving judgment or feeling and measures involving thought or sensorimotor process, $F(1, 33.729) = 0.0113$, p = .916). Effect counts differed significantly between the three types of measures ($\chi^2(2) = 198.330$, p = 8.75E-44), with many more effects for measures involving judgment or feeling (k_{effect} = 293) or behavior (k_{effect} = 143) than measures involving thought or sensorimotor process (k_{effect} = 43), indicating that the measures used in research on cleansing effects have tapped more into psychological and behavioral outcomes than mental or bodily processes.

What Features of Experimental and Statistical Design were Associated with the Strongest Cleansing Effects?

Was the Cleansing Effect Tested for Mediation by Another Measure? 22 cleansing effects were tested for mediation by another measure (vs. 457 were not tested; $\chi^2(1) = 395.042$, p = 6.61E-88), 16 of which were supported (vs. 463 were not tested or not supported; $\chi^2(1) = 417.138$, p = 1.02E-92). Effect sizes were comparable regardless of whether the cleansing effect was tested for mediation (g = 0.354, p = .000193) or not (g = 0.312, p = 3.51E-34), $F(1, 15.252) = 0.329$, p = .575. Effect sizes were also comparable for cleansing effects where mediation was tested and supported (g = 0.406, p = .000678) and for cleansing effects where mediation was not tested or not supported (g = 0.310, p = 3.92E-34), $F(1, 11.526) = 1.229$, p = .290. Likewise, effect sizes were comparable regardless of whether the measure in a cleansing

effect pertained to the mediating variable (g = 0.353, p = .00111) or the dependent variable (g = 0.314, p = 1.04E-35), F(1, 8.273) = 0.284, p = .608.

Did the Study Test Content-Based Specificity Using Another Measure? Of the 398 content-based cleansing effects, 132 came from studies that tested content-based specificity using another measure (vs. 266 from studies that did not; χ^2(1) = 45.116, p = 1.86E-11), 129 of which found support (vs. 269 from studies that did not test or support content-based specificity; χ^2(1) = 49.246, p = 2.26E-12). Effect sizes were comparable regardless of whether the cleansing effects came from studies that tested content-based specificity (g = 0.365, p = 1.69E-11) or not (g = 0.320, p = 4.36E-22), F(1, 62.685) = 0.946, p = .334. But effect sizes were larger for cleansing effects from studies that tested and supported content-based specificity (g = 0.400, p = 3.06E-14) than for cleansing effects from studies that did not test or did not support it (g = 0.308, p = 6.16E-21), F(1, 57.129) = 4.814, p = .032.

Did the Study Test Moderation by Another Factor? 242 cleansing effects came from studies that tested moderation by another factor (vs. 237 from studies that did not; χ^2(1) = 0.0522, p = .819), 173 of which found support (vs. 306 from studies that did not test or support moderation; χ^2(1) = 36.929, p = 1.23E-09). Effect sizes were comparable regardless of whether the cleansing effects came from studies that tested moderation (g = 0.272, p = 2.31E-14) or not (g = 0.342, p = 4.12E-24), F(1, 152.649) = 3.198, p = .076, and whether they came from studies that tested and supported moderation (g = 0.321, p = 3.58E-13) or studies that did not test or did not support it (g = 0.313, p = 1.95E-25), F(1, 100.472) = 0.033, p = .857.

Did the Focal Factor Include Another Condition? 54 cleansing effects included another condition in the focal factor (vs. 425 did not; χ^2(1) = 287.351, p = 1.88E-64). Effect sizes were comparable regardless of whether the cleansing effects included another condition in the focal factor (g = 0.349, p = 6.68E-07) or not (g = 0.309, p = 2.99E-31), F(1, 37.088) = 0.468, p = .498.

Was the Study Preregistered? 450 cleansing effects came from non-preregistered studies (vs. 29 from preregistered studies; χ^2(1) = 370.023, p = 1.85E-82). Effect sizes were significantly larger in non-preregistered studies (g = 0.333, p = 6.66E-39) than in preregistered studies (g = 0.0502, p = .534), F(1, 10.797) = 12.537, p = .005. The statistically non-significant synthesized effect size for preregistered studies raises questions about how to interpret the full body of work on cleansing effects, a theme we will explicate in the **Discussion**.

What Participant Demographics were Associated with the Strongest Cleansing Effects?

Type of Participants. Most cleansing effects involved local or university-related (k_{effect} = 343) or online participants (k_{effect} = 103), and the rest (k_{effect} = 33) involved other types of participants (e.g., healthcare practitioners, political activists, company employees) or did not specify the nature of their participants, χ^2(2) = 331.106, p = 1.26E-72. Effect sizes were comparable between local or university-related participants (g = 0.298, p = 9.00E-23), online participants (g = 0.332, p = 6.31E-13), and other or unspecified participants (g = 0.406, p = .000215), F(2, 38.111) = 0.976, p = .386.

Region of Participants. Effect counts differed significantly between regions (χ^2(3) = 727.121, p = 2.76E-157), with far more cleansing effects based on participants in the West (k_{effect} = 374) than in the East (k_{effect} = 59), Middle East (k_{effect} = 26), and other or unspecified regions (k_{effect} = 20). Effect sizes also differed significantly between them, F(3, 24.020) = 3.822, p = .023. Specifically, effect sizes were smaller in the West (g = 0.288, p = 4.36E-28) than in the East (g = 0.475, p = 5.44E-06; F(1, 24.556) = 5.423, p = .0284) or Middle East (g = 0.462, p = 1.40E-05; F(1, 11.082) = 7.908, p = .0168). No other pairwise comparison was significant (Fs ≤ 2.577, ps ≥ .125).

Gender Ratio of Participants. Effect sizes did not vary significantly as a function of the female percentage of participants in a study, F(1, 31.080) = 1.031, p = .318.

What Report Characteristics were Associated with the Strongest Cleansing Effects?

Year of Report. Effect sizes were significantly smaller in newer reports, $F(1, 35.262) = 6.597$, $p = .015$.

Type of Report. There were many more cleansing effects from peer-reviewed reports ($k_{effect} = 406$, including 366 from peer-reviewed journal articles and 40 from peer-reviewed conference presentations) than from non-peer-reviewed reports ($k_{effect} = 73$, including 35 from unpublished reports, 33 from theses, and 5 from internal conference presentations), $\chi^2(1) = 231.501$, $p = 2.81E\text{-}52$. Effect sizes were significantly larger in peer-reviewed reports ($g = 0.343$, $p = 1.91E\text{-}36$) than in non-peer-reviewed reports ($g = 0.165$, $p = .00456$), $F(1, 39.558) = 9.721$, $p = .003$.

How Report Was Presented by Authors. There were many more cleansing effects in reports presented by its authors as original experiments ($k_{effect} = 401$) than in reports presented by its authors as successful replications ($k_{effect} = 29$) or unsuccessful replications ($k_{effect} = 49$), such that effect counts differed significantly between them ($\chi^2(2) = 548.409$, $p = 8.21E\text{-}120$). Effect sizes also differed significantly between them, $F(2, 13.189) = 60.844$, $p = 2.19E\text{-}07$. Specifically, effect sizes were smaller in reports presented as unsuccessful replications ($g = -0.00283$, $p = .911$) than in reports presented as successful replications ($g = 0.269$, $p = .00452$; $F(1, 7.811) = 19.107$, $p = .00252$) or as original experiments ($g = 0.365$, $p = 8.90E\text{-}39$; $F(1, 28.996) = 130.226$, $p = 3.05E\text{-}12$), but were comparable between reports presented as successful replications and as original experiments, $F(1, 5.743) = 2.541$, $p = .164$.

Did Directionality Interact with Other Moderators of Cleansing Effects?

Table 4 shows the omnibus significance tests for the double-moderator interactions between directionality (i.e., whether an effect demonstrated a psychological consequence or antecedent of cleansing) and each other categorical or continuous moderator. Results based on the RE meta-regression using RVE with SSC are depicted in Figure 3 and briefly outlined below. Further results based on PET and PEESE using RVE with SSC are also available in Table 4 and depicted in Figures S11-S12. For each double-moderator combination, the statistical details for both the simple effects of all levels (comparison against $g = 0$) and the pairwise comparisons between all levels are available in the supplemental material on OSF.

Effect sizes varied significantly as a function of the interactions of directionality with (a) theoretical basis of the cleansing effect (content-based vs. procedure-based), (b) content-based assimilation vs. contrast of the cleansing effect, (c) relation of the psychological variable to morality, (d) moral subdomain of the psychological variable, (e) type of participants, and (f) female percentage of participants. A descriptive summary of these interactions is available in Supplemental Material C; our conceptual interpretation of them is provided in the **Discussion**. No other double-moderator interactions with directionality were significant.

Discussion

Early interests in the psychology of physical cleansing focused on its antecedents from health and clinical perspectives. For example, people in various naturalistic contexts (e.g., public bathroom, hospital) were more likely to clean their hands if they had (vs. had not) seen an observer (Munger & Harris, 1989) or if they knew (vs. did not know) they were being observed (Eckmanns et al., 2006; Edwards et al., 2002; Maury et al., 2006). Women were more likely to exhibit spontaneous washing behavior if they had been prompted to imagine receiving a non-consensual (vs. consensual) kiss (Fairbrother et al., 2005).

A new wave of interest emerged in the mid-2000s, with a different empirical focus and methodological style, thanks to a couple of seminal papers that examined the mental associations of cleansing with morality from the perspective of conceptual metaphor (Schnall et

al., 2008; Zhong & Liljenquist, 2006). Research on cleansing has since gained momentum in social, personality, and consumer psychology, coinciding with the popularity of research on morality and disgust, and broadened substantially to a diverse array of psychological consequences and antecedents, from morality, politics, and religion to decision making, risk taking, and attitude change.

Much of this development unfolded against the backdrop of the replicability crisis and open science movement that started in 2011. With the reporting of some failed replications of the original findings (Earp et al., 2014; Fayard et al., 2009; Gámez et al., 2011; D. J. Johnson et al., 2014a), interests in cleansing effects (and more broadly, conceptual metaphors) took on a different flavor. A focal meta-analysis (k_{effect} = 15) examining "the effects of an immoral versus moral prime on cleansing-related preferences or behaviors" (Siev et al., 2018, p. 2) estimated a small weighted mean effect size of g = 0.17, p = .013, 95% CI [0.04, 0.31] if it included all 15 effects (from both original and replication studies). But if it excluded the 3 original effects and only included the other 12 effects, it estimated a non-significant weighted mean effect size of g = 0.07, p = .207, 95% CI [-0.04, 0.19]. These findings about the specific "immorality → cleansing" effect reinforced concerns about the replicability of experimental effects on the link between cleansing and morality in general.

Such concerns stood in stark contrast to the typical emphasis on significant findings in narrative reviews of cleansing effects (S. W. S. Lee & Schwarz, 2011, 2016, 2021; West & Zhong, 2015; Yan, 2011). In our view, they highlighted the need for and potential utility of a comprehensive quantitative synthesis of all relevant empirical evidence. The present meta-analysis sought to do so by casting a wide net (129 unique records, 230 experiments, and 551 effects based on 42,793 participants), covering all experimental findings related to cleansing regardless of what theoretical traditions they were couched in (e.g., social presence, conceptual metaphor). We identified and removed outliers (e.g., leave-one-out diagnostics), assessed and corrected for publication bias (e.g., parameter selection model), compared results from non-preregistered vs. preregistered studies, and compared results from original experiments vs. replications (both successful and unsuccessful ones). We also analyzed a variety of other moderators (theoretical, methodological, participant demographics, report characteristics). These attributes, and the basic goal of meta-analysis, shift the focus away from singling out individual studies—be they original experiments or replications, preregistered or not—towards taking the full population of experiments into account and extracting reliable patterns and nuances from them (B. T. Johnson, 2021).

What Have We Learned from this Meta-Analysis?

Outliers, Publication Bias, and Replicability

We made efforts to assess and correct for outliers (using leave-one-out diagnostics and sensitivity analysis) and publication bias (using eight methods). Results suggested that the observed pattern of meta-analytic findings was unlikely to be mere artifacts. Even after the most severe corrections, synthesized effects were significant though small in size (Table 3).

A reasonable critique, however, could be "garbage in, garbage out." If many of the experiments included in the dataset were products of questionable research practices (e.g., *p*-hacking, selective reporting; John et al., 2012; Simmons et al., 2011), then whatever meta-analytic conclusions are drawn would have to be discounted. This is precisely the concern that *p*-uniform (van Assen et al., 2015) and *p*-uniform* (van Aert & van Assen, 2018) were designed to address. Both approaches showed significant synthesized effects, gs = 0.277 and 0.220, respectively (Table 3, Methods 5a-5b).

This rosy conclusion may still be undermined by caveats against the performance of *p*-uniform (and *p*-uniform*) under research conditions such as large heterogeneity, publication bias, and *p*-hacking (van Aert et al., 2016). Under these conditions, parameter selection model

(PSM) is the preferred approach (McShane et al., 2016). When we implemented PSM, regardless of whether we used three or four parameters, the synthesized effect remained significant, though again in the small range, *g*s = 0.209 and 0.164, respectively (Table 3, Methods 6a-6b).

In spite of the overall evidential value suggested by our application of state-of-the-art bias correction methods, additional methodological concerns (e.g., replicability, robustness, preregistration, power) have been voiced about this area of work, echoing the broader movement in psychology and other sciences (Nelson et al., 2018; Nosek et al., 2022; Open Science Collaboration, 2015). Six observations from the present meta-analysis are relevant to these concerns. Taken together, they suggest that much of the evidence for cleansing effects came from studies conducted without upholding all of the contemporary standards of scientific rigor.

First, the majority of the meta-analyzed cleansing effects (450 out of 479) came from non-preregistered studies, where the synthesized effect size was g = 0.333 (Figure 2). Few effects (29 out of 479) came from preregistered studies, where the synthesized effect size was g = 0.050. The discrepancy highlights the importance of conducting preregistered studies in future work. It also signals a need for downward adjustment to the confidence one may ascribe to conclusions drawn from the present set of mostly non-preregistered studies.

Second, among the meta-analyzed cleansing effects in reports presented by their authors as original experiments (k_{effect} = 401), only 6 effects came from preregistered studies, all in a single report (S. W. S. Lee et al., 2023). Among the meta-analyzed effects in reports presented by their authors as unsuccessful replications (k_{effect} = 49), 23 effects came from preregistered studies (the details of which are available in Supplemental Material D) and 26 effects came from non-preregistered studies. Successful replications also exist, including 29 of the meta-analyzed effects plus 3 effects excluded from the meta-analysis because of their large effect sizes (*g*s > 1.0). All 32 effects came from non-preregistered studies.

Third, some of the experiments presented as replications contained methodological deviations from the original experiments, such as different procedures (Camerer et al., 2018), different manipulations (Earp et al., 2014), and different cultural samples (Gámez et al., 2011). In the presence of such deviations, the replications may be critiqued as non-exact, but they may still serve the function of calibrating the generalizability of the original findings.

Fourth, there were both successful and unsuccessful replications of the same original experiment. For example, three replications (D. J. Johnson et al., 2014a, 2014b) did not successfully replicate the original findings by Schnall et al. (2008), but two other direct replications (Arbesfeld et al., 2014; Besman et al., 2013) and three extended replications (J. L. Huang, 2014) did successfully replicate the original findings. It is noteworthy that the published report of unsuccessful replications (D. J. Johnson et al., 2014a) has received considerably more attention (161 citations on Google Scholar as of September 27, 2023) than the published report of successful replications with extensions (J. L. Huang, 2014; 29 citations), despite the latter's larger sample sizes. The difference in attention may be unsurprising, considering that our field has been grappling with replicability concerns. At the same time, it signals a possible difference in salience of successful and unsuccessful replications and suggests that caution is warranted in taking stock of both kinds of replications, particularly when they pertain to the same original experiment.

Fifth, a more general pattern is that some original experiments have been successfully replicated and others unsuccessfully replicated. Because successful and unsuccessful replications pertained to different original experiments that asked different specific research questions, fine-grained comparisons and interpretations are difficult. With this caveat in mind, we offer a brief summary of the broad-stroke differences between successful and unsuccessful replications, in hopes of sketching the landscape to the best of our knowledge: Successful replications have only been demonstrated in psychological consequences of cleansing (k_{effect} = 32), regardless of whether they were content-based and directly related to morality (k_{effect} = 26)

or procedure-based and unrelated to morality (k_{effect} = 6); all of the content-based consequences were assimilative ("cleansing → more moral"). Unsuccessful replications have been demonstrated in both psychological consequences (k_{effect} = 29) and antecedents (k_{effect} = 20) of cleansing. Most were content-based (k_{effect} = 39) and related to morality (either directly [k_{effect} = 36] or indirectly [k_{effect} = 3]), about half of which were assimilative (k_{effect} = 17) and half contrastive (k_{effect} = 22). Some were procedure-based and unrelated to morality (k_{effect} = 10).

Sixth and finally, the typical experiment was underpowered to detect the observed effect size. By "typical experiment," we assume a between-participant experimental design of one factor with two conditions. (Of course, actual experiments involved other designs—such as within-participant, more than one factor, and more than two conditions in a factor—which would have different power than the "typical experiment.") As shown in Table 6, we conducted power analysis to estimate the statistical power of an experiment with the median sample size and the synthesized effect size (based on the RE or FE model) to show a significant effect (alpha = .05). We did this for all experiments, or original experiments only, or all replications, or successful replications only, or unsuccessful replications only. Regardless of which set of experiments we looked at, power consistently fell below the standard recommendation of .80. Two original experiments and one unsuccessful replication had unusually large sample sizes (*N*s = 7,001, 4,142, and 2,808); excluding them did not change our conclusions. These observations suggest that to reliably ascertain the conditions under which cleansing effects are observable and replicable, future experiments will need to attain higher power by increasing effect size (e.g., using more effective manipulations), sample size, or both (Cohen, 2013). In hopes of promoting such endeavours, we provide the sample sizes required to detect various benchmark effect sizes with power = .80 and alpha = .05 in Table 6 (rightmost column).

How Do Cleansing Effects Operate?

In addition to critically assessing the methodological weaknesses of the extant body of research on cleansing effects overall, we also consider it our responsibility as meta-analysts to discern empirical patterns and nuances that can inform theoretical development and guide future work. Whether these findings will stand the test of replicability remains to be seen. But to ignore them is, in our view, to ignore the opportunity to learn something potentially useful (B. T. Johnson, 2021). Therefore, we glean all results of the moderator analyses to inform our understanding of the theoretical properties of cleansing effects. In a nutshell, cleansing effects emerge across psychological domains and operate bidirectionally (Figure 4).

Scope. Cleansing effects are not limited to the moral realm, nor are they limited to disgusting contexts. Synthesized effect size estimates were similar regardless of whether the psychological domain was directly related, indirectly related, or unrelated to morality (Table 4 and Figure 2). Among effects directly related to morality, effect size estimates were comparable across moral subdomains. Among effects *not* directly related to morality, effect size estimates were comparable across social vs. non-social domains. Many of the psychological variables in the meta-analyzed experiments involved no disgusting elements at all (e.g., fairness, empathy, postdecisional dissonance, luck).

The domain-general nature of the psychological consequences and antecedents of cleansing is most consistent with the theoretical perspective of cleansing as a grounded procedure of separation (S. W. S. Lee & Schwarz, 2021). This perspective postulates that physical acts of cleansing oneself can serve as sensorimotor experiences that ground mental procedures of separation, which can be applied across domains to confer a sense of psychological separation between an event and oneself. The mechanism of separation predicts that cleansing has domain-general psychological consequences because it functions as a process that reduces the residual influence of a prior experience by mentally separating it from the present self. Indeed, the meta-analytic data include evidence that wiping or washing one's hands could reduce the residual influence of a recent product choice on one's subsequent product

evaluation (S. W. S. Lee & Schwarz, 2010b; also De Los Reyes et al., 2012), reduce the residual influence of a recent academic failure on one's pessimism about one's future performance (Kaspar, 2012), reduce the residual influence of a recent product endowment on one's subsequent attachment to it (Florack et al., 2014), reduce the residual influence of a recent lucky or unlucky streak on one's subsequent betting behavior (A. J. Xu et al., 2012; also Moscatiello & Nagel, 2014), and more.

The mechanism of separation also predicts that cleansing has domain-general psychological antecedents because negative psychological experiences in a variety of domains can trigger the desire for separation, which can be reflected in higher accessibility of cleansing-related thoughts and stronger desires for cleansing behavior or cleansing products. For example, experimental evidence shows that imagining a non-consensual kiss led women to have a stronger urge to clean themselves (Bilekli & Inozu, 2018), that using race-tainted evidence in a legal case increased lawyers' and law students' tendency to choose a hand gel rather than a pen as a free gift (Bilz, 2012), and that smelling a shirt belonging to an outgroup member increased participants' usage of a hand sanitizer (Reicher et al., 2016). The domain-general nature of the psychological antecedents and consequences of physical cleansing goes beyond the scope of morality-based or disgust-based accounts of cleansing effects.

Directionality. Cleansing manipulations can influence psychological measures, and psychological manipulations can influence cleansing-related measures. Both directions of cleansing effects exhibited comparable effect sizes overall (Figures 2 and 4). The bidirectional nature of cleansing effects is consistent with the frequently observed bidirectional links between mental and bodily processes (S. W. S. Lee & Schwarz, 2012).

Directionality did interact with several theoretical moderators (Table 5). These included theoretical basis of the cleansing effect, whether it was about content-based assimilation or contrast, relation of the psychological variable to morality, and moral subdomain of the psychological variable (Figure 3 and Supplemental Material C). We explore each in turn.

Psychological consequences of cleansing showed larger effect sizes for content-based than procedure-based effects, but psychological antecedents of cleansing showed comparable effect sizes for content-based and procedure-based effects. The synthesized effect sizes in all four categories were significant. This pattern suggests that cleansing effects can operate on the bases of both content and procedure. Recall that for a cleansing effect to be considered procedure-based, its psychological variable must have no shared content with physical cleansing. If there was any content overlap, the cleansing effect was considered content-based—even if procedures (e.g., separating past from present) might also be at work. (We cannot think of any feasible way, in principle or in practice, to guarantee that content-based effects were devoid of procedures.) That means procedure-based effects had to be driven by procedure alone, whereas content-based effects could be driven by content alone or by both content and procedure. This might be the reason that effect sizes were larger for content-based psychological consequences of cleansing (where both content and procedure could be at work) than procedure-based ones (where only procedure could be at work).

As far as content-based effects are concerned, within the category of content-based contrast, psychological consequences of cleansing (e.g., cleansing oneself decreases subsequent volunteering behavior) showed larger effect sizes than did psychological antecedents of cleansing (e.g., recalling an unethical behavior increases one's desire for cleansing products). It appears that cleansing absolves guilt more than guilt prompts cleansing. But within the category of content-based assimilation, psychological consequences of cleansing (e.g., feeling clean increases donating behavior) and psychological antecedents of cleansing (e.g., unscrambling religion-related words increases the number of cleansing-related words completed) showed comparable effect sizes. Again, the synthesized effect sizes in all four categories were significant. These observations indicate that the general pattern of bidirectionality is underlain by specific

nuances of variability. In particular, they highlight the guilt-absolving power of cleansing (contrast) and the conceptual links between clean and moral (assimilation).

Turning to the interaction of directionality with the nature of the psychological variable, psychological consequences of cleansing showed larger effect sizes for psychological variables directly related to morality (e.g., cleansing influences moral judgment) than for those unrelated to morality (e.g., cleansing influences information processing). Effect sizes for psychological consequences indirectly related to morality (e.g., cleansing influences healthy/unhealthy eating) were descriptively close to those directly related to morality, though not significantly different from those unrelated to morality due to relatively low effect counts (and thus larger standard errors). In contrast, psychological antecedents of cleansing showed larger effect sizes for psychological variables indirectly related to morality (e.g., prompting empathy increases hand-hygiene behavior) than for those directly related to morality (e.g., copying a story of unethical behavior increases desirability of cleansing products). This difference, however, should be interpreted with caution due to the very low effect count for psychological antecedents indirectly related to morality. Effect sizes for psychological antecedents unrelated to morality (e.g., social presence of others increases cleansing behaviors in the bathroom) sat between those indirectly related to morality and those directly related to morality, with no significant difference from either.

Finally, among psychological variables directly related to morality, psychological consequences of cleansing showed larger effect sizes for fairness/cheating than for care/harm and sanctity/degradation, though the synthesized effect sizes in all categories were significant. In contrast, psychological antecedents of cleansing showed smaller effect sizes for fairness/cheating than for sanctity/degradation and other subdomains, mixed subdomains, or morality in general; the synthesized effect sizes were significant for sanctity/degradation and "other/mixed/general," but not significant for fairness/cheating and care/harm. These patterns may be interpreted through several theoretical lenses, which characterize the operation of cleansing effects via different mechanisms.

As noted in the preceding subsection (**Scope**), the grounded procedure perspective predicts domain-general psychological consequences of cleansing, because it functions as a procedure that reduces the residual influence of a prior experience by mentally separating it from the present self (S. W. S. Lee & Schwarz, 2021). This perspective also predicts domain-general psychological antecedents of cleansing, because negative psychological experiences across domains can trigger desires for separation and thus cleansing-related thoughts and desires. If this psychological mechanism is at work, psychological experiences that feel more negative should more powerfully instigate cleansing-related outcomes. This property may underlie the meta-analytic finding that psychological antecedents of cleansing showed particularly large effect sizes if they were about sanctity/degradation (k_{effect} = 61), almost all of which involved sexual content (k_{effect} = 60; only 1 involved non-sexual content). The specific psychological manipulations related to sexual sanctity/degradation (e.g., imagining a non-consensual kiss) in the meta-analyzed set of experiments might have been more potent in negative valence than, for example, those related to fairness/cheating (e.g., receiving an unfair offer). The potent negativity of sanctity/degradation also implies that it is harder to be removed by a cleansing manipulation, such that the effects of cleansing on sanctity/degradation should be weaker. Indeed, psychological consequences of cleansing showed smaller effect sizes for sanctity/degradation than for fairness/cheating.

An alternative interpretation is that sanctity/degradation violations are the kind of immorality most prototypically associated with the physical domain of cleanliness and purity (Graham et al., 2011, 2013), such that they trigger a particularly strong domain-specific mapping. Relatedly, it is possible that sanctity/degradation violations involve physical disgust more so than other moral violations do (Rozin et al., 2008; Rozin & Fallon, 1987), such that people are particularly motivated to engage in decontamination behaviors such as cleansing.

These mechanisms—disgust, domain-specific mapping, and grounded procedure of separation—are not mutually exclusive. They can co-occur in real life and remain to be teased apart in research (see ***What Aspects of Cleansing Effects Are Unexplored or Underexplored?***).

What Methodological Aspects of Cleansing Experiments Matter?

Operationalizations of Manipulation. Overall, different types of manipulation had similar effect sizes, regardless of whether they involved actual experience, imagined or recalled experience, or conceptual activation (Table 4). This was true among both psychological consequences and antecedents of cleansing (Table 5). These general observations are grounded in a rich variety of specific manipulations. We provide a summary of them in Supplemental Material E both to convey the methodological richness and to help researchers pick the most applicable manipulations for their future work.

Operationalizations of Measure. Different types of measure had different effect sizes. Measures of behavior showed significantly larger effect sizes than measures of judgment or feeling (Figure 2), a pattern that was evident among both psychological consequences and antecedents of cleansing (Figure 3). Measures of behavior also showed larger effect sizes than measures of thought or sensorimotor process, though this pattern was evident only among psychological consequences of cleansing, not among psychological antecedents of cleansing. Essentially, cleansing manipulations exerted stronger effects on behavioral measures than on mental measures of psychological outcomes. Likewise, manipulations of psychological variables exerted stronger effects on behavioral measures than on judgment- and feeling-based measures of cleansing-related outcomes.

The existence of strong cleansing effects on behavioral measures is noteworthy against the scientific backdrop that social and personality psychology over the years has focused less and less on actual behaviors and more and more on "introspective self-reports, hypothetical scenarios, and questionnaire ratings" (Baumeister et al., 2007, p. 396). On the one hand, in terms of effect counts, we do see more cleansing effects on measures of judgment or feeling than on measures of behavior (Figure S7). On the other hand, in terms of effect sizes, we see the strongest cleansing effects on behavioral measures of both psychological outcomes and cleansing-related outcomes (Figure 3).

Indeed, psychological outcomes of cleansing have been measured with a wide spectrum of actual behaviors, including social and economic behaviors, choice behaviors in consumer contexts, and physical behaviors, as summarized in Supplemental Material F. Cleansing-related outcomes have also been measured using a number of actual behaviors. All of these suggest that the psychology of cleansing is more than a "science of self-reports and finger movements" (Baumeister et al., 2007, p. 396). It involves overt behaviors as well as covert mental processes.

Features of Experimental and Statistical Design. Although cleansing effects were larger for experiments that supported content-based specificity using another measure than for experiments that did not support it, other features of experimental and statistical design were not significant moderators (Table 4). Specifically, cleansing effects were similar in size for dependent or mediating variables, for experiments that did or did not test mediation, that did or did not support such mediation, that did or did not test moderation by another factor, that did or did not support such moderation, and that did or did not include another condition in the focal factor (Figure 2).

Participant Demographics. Effect sizes were smaller in the West than in the East or Middle East (Figure 2), suggesting cultural variation in cleansing effects. Meanwhile, far more cleansing effects were based on participants in the West than elsewhere (Figure S7). This may be unsurprising but does echo broader calls for the need to globalize psychological research because many psychological effects are known or expected to differ across world regions (Henrich et al., 2010; Thalmayer et al., 2021). Another possible reason for this West-heavy

distribution of effects is that the database we used to search for relevant studies (APA PsycInfo) and the other sources of our studies (conference proceedings, PsychFileDrawer.org, recent large-scale replication projects, and listserv requests) may have over-emphasized research in the West and scholarly work written in English (see ***Limitations of Our Choice of Evidence Base***).[15]

Although cleansing effects overall did not vary in size between different types of participants or as a function of female percentage of participants, these moderators interacted with directionality (Table 5). Specifically, psychological consequences of cleansing showed smaller effect sizes in local, university-related, or online participants than in other or unspecified types of participants (though this last category contained only eight effects); psychological antecedents of cleansing showed similar effect sizes across types of participants (Figure 3). A higher female percentage of participants was associated with a larger increase in effect sizes among psychological consequences of cleansing than among psychological antecedents of cleansing. The heterogeneity of cleansing effects across participant demographics suggests the need to be cautious about generalizing conclusions across hitherto unexamined populations.

Report Characteristics. Finally, cleansing effects varied in size depending on several report characteristics (Table 4). Effect sizes were larger in peer-reviewed than non-peer-reviewed reports (Figure 2), indicating a probable reporting bias (B. T. Johnson & Hennessy, 2019). Unsurprisingly, effect sizes were larger in reports presented by their authors as original experiments or successful replications than as unsuccessful replications (Figure 2). Effect sizes were also larger in older reports (Table 4). It may reflect changing norms of research practices and resulting findings in our field. It may also reflect other differences between studies conducted in earlier vs. more recent years, such as experimental technique, topical focus, and research goal.[16]

What Research Gaps and Questions are Highlighted by this Meta-Analysis?

Throughout our process of examining synthesized effects (both overall and across moderators), we focused on effect sizes and excluded outliers so as to obtain estimates that were as unbiased as possible (see **Outliers and Publication Bias**). But for the purpose of identifying research gaps, we will focus on effect counts and consider all effects (Table S1, first column of results) in order to depict the most comprehensive landscape of what has been studied and what remains to be explored.

What Are the Methodological Limitations of Cleansing Effects Thus Far?

Research on cleansing effects thus far has been more outcome-oriented than process-oriented. There have been far more effects that measured psychological and behavioral outcomes (k_{effect} = 338 on judgment or feeling; k_{effect} = 167 on behavior) than those that measured mental or bodily processes (k_{effect} = 46, which included 22 on concept accessibility, 17 on cognitive process, and 7 on sensorimotor process). Mediation tests have been scarce (25 cleansing effects were tested for mediation by another measure, 526 were not). Process insights, however, can also be gained by moderation (Pirlott & MacKinnon, 2016; Spencer et al., 2005). Moderation tests were more common than mediation tests. 272 of the 551 cleansing effects (i.e.,

[15] We thank the Editor for this suggestion.

[16] In addition, the last few years have witnessed the global pandemic of COVID-19. One might expect cleansing effects to be smaller in size during the pandemic because cleansing became particularly tied to disease-related concerns and thus less likely to trigger other psychological associations. Or one might expect cleansing effects to be larger in size during the pandemic because of its general salience. These possibilities were difficult to assess in the present meta-analysis because only a small number of studies were conducted during the pandemic.

49.4%) came from experiments that tested moderation by another factor, 200 (i.e., 73.5%) of which found support.

In terms of discriminant validity, on the side of dependent variables, 141 of the 459 content-based cleansing effects (i.e., 30.7%) came from experiments that tested and supported content-based specificity using another measure. On the side of independent variables, only 72 of the 551 cleansing effects (13.1%) included more than two conditions in the focal factor. These relatively low percentages indicate that future work would benefit from experimental designs that provide informative evidence for the extent to which cleansing effects are distinct from related or more general effects (e.g., affective valence).

Sample diversity is another weakness of the meta-analyzed dataset. As is typical in psychology research (Henrich et al., 2010; Thalmayer et al., 2021), the vast majority of cleansing effects involved local or university-related (k_{effect} = 396, i.e., 71.9%) or online participants (k_{effect} = 107, i.e., 19.4%). The dominant region of focus is the West (k_{effect} = 423, or 76.8%). These represent a thin and atypical slice of humanity. Generalizability to other populations remains largely unknown.

In addition to addressing these methodological limitations, future research on cleansing would benefit from careful consideration of effect size and statistical power. Results from the present meta-analysis suggest that although cleansing effects appear robust across many moderators (e.g., psychological consequences vs. antecedents of cleansing, content-based vs. procedure-based effects), there are cases where effect sizes varied as a function of theoretical focus (e.g., different subdomains of cleansing) or methodological choice (e.g., measures of behavior vs. judgment or feeling). Recognizing that the typical experiment on cleansing effects is underpowered, we join the chorus in our field in recommending that future work be informed by power analysis, leveraging the effect size estimates provided by the present meta-analysis to determine the sample size required to detect a particular cleansing effect of interest with sufficient power. Table 6 provides relevant estimates for a simple experiment with two conditions between participants. For more complex predictions (e.g., ordinal interaction), power analyses will be necessary and considerably larger sample sizes should be expected (Giner-Sorolla, 2018; Lakens, 2020; Simonsohn, 2014).

In thinking about statistical power, a common recognition is that all else being equal, within-participant designs afford greater statistical power than between-participant designs. But when it comes to examining the psychological consequences of cleansing, it is practically challenging to manipulate cleansing in a within-participant design because it would require the same participant to experience both cleansing and non-cleansing under otherwise identical conditions and respond to the same measures more than once, which increases the risks of suspicion, reactance, hypothesis awareness, and demand effects. Presumably for these reasons, cleansing manipulations overwhelmingly relied on between-participant designs. Of the various types of cleansing manipulations, those involving actual cleansing experience require in-person experimentation, for which large sample sizes are often difficult to attain. In contrast, those involving imagined cleansing experience or merely conceptual activation of cleansing can be implemented in online settings, for which large sample sizes are easier to attain. But online settings come with other limitations, such as difficulty of observing actual behavioral outcomes and difficulty of implementing longer experimental procedures (due to online participants' limited attention span). In short, there is no perfect design. Resources will need to be abundant or compromises will need to be made in the pursuit of higher statistical power in future tests of cleansing effects.

No matter what design is adopted, we recommend that future studies be preregistered. The preregistration should clearly specify the study design, power analysis, sample size, stopping rules in data collection, steps of data preprocessing, inclusion/exclusion criteria, and statistical analyses. Another important aspect that should be specified in the preregistration is the precise direction of the hypothesized main effect (e.g., an assimilation effect where cleansing

increases moral behavior; a contrast effect where cleansing increases immoral behavior) or the precise shape of the hypothesized interaction effect (e.g., an attenuated interaction where cleansing increases moral behavior in one condition but has no influence in another condition; a reversed interaction where cleansing increases moral behavior in one condition but increases immoral behavior in another condition). These divergent patterns of results have been found in the extant literature of mostly non-preregistered experiments, presumably lending support to the broad notion of cleansing effects. But clearly, they are not one and the same. By specifying the precise direction or shape of the hypothesized effect, future studies will provide diagnostic evidence that informs critical theoretical assumptions (e.g., under what conditions cleansing should be expected to produce assimilation vs. contrast effects; see **Content-Based Assimilation vs. Contrast** under ***What Aspects of Cleansing Effects Are Unexplored or Underexplored?***). More broadly, the most informative experimental designs should not only incorporate all of the methodological considerations above (process evidence, discriminant validity, sample diversity, effect size, statistical power, sample size, preregistration), but also tackle the most meaningful theoretical questions, especially those that remain unexplored or underexplored, as elaborated below.

What Aspects of Cleansing Effects Are Unexplored or Underexplored?

Domains of Psychological Experience. Procedure-based cleansing effects have received considerably less empirical attention (k_{effect} = 91) than content-based ones (k_{effect} = 460), in both original experiments (75 procedure-based, 395 content-based) and replications (16 procedure-based, 65 content-based), despite their comparable effect sizes (Figure 2). Relatedly, in psychological domains directly and indirectly related to morality, similar amounts of attention have been paid to the psychological consequences (k_{effect} = 245) and antecedents (k_{effect} = 189) of cleansing, but in psychological domains *unrelated* to morality, much more attention has been paid to the psychological consequences (k_{effect} = 82) than antecedents (k_{effect} = 35) of cleansing, despite their similar effect sizes (Figure 3). In other words, as we go beyond the realm of morality, we know much more about what outcomes are influenced by cleansing than what influences cleansing-related outcomes. This is an important empirical limitation because although the domain-generality of psychological consequences and antecedents of cleansing is compatible with the perspective of cleansing as a grounded procedure of separation (S. W. S. Lee & Schwarz, 2021), there is not nearly as much evidence for procedure-based effects as for content-based ones, and the evidence thus far in non-moral domains has been much more substantial for one causal direction than the other. Based on our meta-analytic results, we expect future research to find that both psychological consequences and antecedents of cleansing are observable across domains, be they directly related, indirectly related, or unrelated to morality (Figure 4).

Within the moral domain, particularly large effect sizes were found if the psychological antecedents of cleansing were about sanctity-degradation (k_{effect} = 61, of which 60 were sexual and 1 was non-sexual). As mentioned earlier (**Directionality** under ***How Do Cleansing Effects Operate?***), this pattern may be interpreted through several non-mutually exclusive mechanisms, including (1) grounded procedure of separation (sanctity-degradation violations being more negative than others), (2) domain-specific mapping (sanctity-degradation violations being prototypically associated with cleanliness), and (3) disgust (sanctity-degradation violations being particularly disgusting). To tease apart these mechanisms, future experiments should control for subjective valence intensity in comparing effects across moral subdomains. If, once subjective valence intensity is held constant, sanctity/degradation violations no longer produce stronger cleansing effects than do other moral violations, it would favor the theoretical interpretation that grounded procedure of separation is the mechanism at work. But if, despite equal subjective valence intensity, sanctity/degradation violations still produce stronger cleansing effects than do other moral violations, it would favor the mechanisms of domain-

specific mapping and disgust. Direct evidence bearing on these two mechanisms may also be obtained by measuring mental accessibility of cleansing-related concepts and activation of various components of disgust (e.g., subjective feelings, facial expressions, other physiological changes).

Controlling for subjective valence intensity, by statistics or by design, will be useful for comparing antecedents of cleansing not only across different subdomains of morality, but also across different psychological domains (e.g., does being socially rejected elicit stronger cleansing desires than being financially unlucky?). Furthermore, it will be useful for comparing negative vs. positive experiences (e.g., do losses or immoral acts motivate people to engage in cleansing more strongly than gains or moral acts motivate people to avoid cleansing?). Cleansing experiments thus far tended to instantiate a single negative experience (e.g., lying) versus a single positive experience (e.g., truth-telling) without measuring, let alone equalizing, their subjective intensity. Instantiating gradations of valence will strengthen future experiments in terms of their empirical nuances and inferences for theoretical understanding of what motivates cleansing at what levels of strength. Drawing on nuances revealed by the present meta-analysis, we expect that more intense negative experiences will more strongly evoke cleansing-related thoughts, feelings, and behaviors, but cleansing will have weaker effects on more intense negative experiences because it is harder to separate such experiences from oneself. In other words, the intensity of negative experiences is expected to have a positive relationship with the effect sizes of psychological antecedents of cleansing, but a negative relationship with the effect sizes of psychological consequences of cleansing.

Forms of Physical Cleansing. Cleansing behavior takes many forms in daily life. For example, people may cleanse different body parts (e.g., hands, mouth, face, hair, feet, whole body) or external objects (e.g., dishes, clothes, table, floor). Experimental work, however, has focused largely on manipulations and measures of washing/wiping one's hands, rinsing one's mouth, or being clean oneself (e.g., imagination of having fresh breath and being well-groomed). Exceptions were few and far between, such as manipulating other forms of cleansing (e.g., wiping a board, wiping one's face; Körner & Strack, 2019; S. W. S. Lee et al., 2015; Li et al., 2017; Lobel et al., 2015; Tang et al., 2017) and measuring desires for or actions of cleansing other body parts, the body in general (Chan, 2019; Elliott & Radomsky, 2009; Kwok, 2010; S. W. S. Lee et al., 2015), or external objects (D'Olimpio & Mancini, 2014; Gilchrist & Schnall, 2018; Ottaviani et al., 2018).

Why are cleansing effects with different body parts or with external objects worth exploring? Because different body parts can be associated with different cultural meanings (e.g., the concept of face in East Asian cultures; S. W. S. Lee et al., 2015). Different body parts can also have different degrees of situational salience in different experimental contexts (e.g., conveying an unethical message by typing it on email vs. saying it on voice mail; S. W. S. Lee & Schwarz, 2010a). Both types of factors—chronically different cultural meanings and momentarily different situational salience—have been shown to moderate cleansing effects. Cleansing external objects may also produce different effects than cleansing oneself (e.g., Körner & Strack, 2019). Based on the experimental evidence available to date, we expect psychological consequences and antecedents of cleansing to be stronger when cleansing involves the same body part or the same target as in the psychological experience (e.g., receiving an unwanted kiss and rinsing one's mouth, seeing an immoral scene and wiping the screen).

Within the context of cleansing oneself, lay people have this notion of *inner* cleansing (i.e., cleansing one's body from within), different from the outer cleansing (i.e., cleansing one's body surface) examined in all of the meta-analyzed experiments. Popular interest in the psychological power of inner cleansing may be exemplified by the blossoming variety of "detox & cleanse" products (https://amzn.to/2PzThKQ) and recipes of "juice cleanse" (Valliant, 2012) and "clean eating" (https://amzn.to/2Pyb7hm). What kinds of individuals are most likely to believe in and devote money, time, and efforts to inner cleansing? What are the psychological

antecedents of such beliefs and behaviors? Do people actually feel better afterwards? If so, what characteristics predict who feels better and who does not? Among those who feel better, is it a placebo effect or not? Does inner cleansing, by purifying one's "inner essence," produce even stronger effects than outer cleansing? If so, is it especially true for people who are high on psychological essentialism (Gelman, 2004; Medin & Ortony, 1989) or who embrace an ethics of covert thoughts and intentions ("Gesinnungsethik"; Weber, 1919) as opposed to an ethics of overt behaviors and consequences ("Verantwortungsethik")? These questions have not been addressed at all by the existing body of experimental research on cleansing.

Beyond cleansing oneself and one's environment, what are people's normative expectations about others' need to stay clean? What cultural forces and situational factors shape these expectations? In many cultural contexts where power differentials exist, it appears that the subordinate are expected to keep themselves clean more than the superordinate are. In Biblical times, a person was supposed to cleanse themselves before seeing the priest, and both were supposed to cleanse themselves before God. Some societies prohibited women during menstruation from entering social spaces. Closer to home, anecdotal observations suggest that graduate school interviewees are expected to be well-groomed more than their interviewers are. Further examples abound in other contexts. Descriptive and injunctive norms of cleanliness, as well as their social underpinnings and cultural sanctions, await systematic investigation.[17]

Content-Based Assimilation vs. Contrast. What types of cleansing effects in the literature are more likely to be about content-based assimilation or contrast? Table S6 presents a detailed breakdown of assimilation and contrast effects as a function of directionality, operationalization of manipulation, and operationalization of measure. A summary is available in Supplemental Material G.

The distribution of assimilation and contrast across different types of cleansing effects may be interpreted through different mechanisms, but direct tests of these mechanisms are lacking. The exact conditions under which psychological consequences and antecedents of cleansing are more likely to show assimilation than contrast, or vice versa, remain to be specified.[18] Juxtaposition of a few existing experiments does offer some clues.

Specifically, the effect of a cleansing manipulation may depend on whether it follows a prior manipulation and what that was. In an experiment where participants washed their hands (vs. not) after watching a disgusting movie clip, the cleansing manipulation decreased participants' harshness in judging others' moral transgressions described in vignettes (i.e., an assimilation effect; Schnall et al., 2008, Experiment 2)[19], presumably by washing away some of the residual disgusting feelings that would have borne on the moral judgments. But in another experiment where participants simply washed their hands (vs. not) without any prior manipulation, the cleansing manipulation *increased* participants' harshness in judging social issues (e.g., abortion, recreational drug use) as immoral (i.e., a contrast effect; Zhong et al., 2010, Experiment 1), presumably because in the absence of any specific salient experience, the cleansing manipulation conferred a sense of cleanliness that bore on the self, leading participants to see themselves as physically and morally cleaner, hence their stronger disapproval of morally dirty behaviors (Experiment 3). Across these two experiments, the manipulation was the same, but it took place in different contexts, so it had different psychological meanings and produced different effects.

---

[17] We thank Mark Landau for these suggestions.

[18] An empty cell in Table S6 may be empty because its corresponding effects are theoretically unlikely to occur, or because there happens to be no empirical attention to it yet.

[19] Note that the effect observed in this experiment (Experiment 2) has not been subject to tests of replicability, unlike the effect observed in the other experiment in the same report (Experiment 1; Schnall et al., 2008).

Compatible with this context-sensitive view of cleansing effects were the results of a 2 × 2 experiment (Khan & Grisham, 2018). Participants first wrote about an experience that made them feel either immoral or emotionally neutral, and then were asked to either wipe their hands clean or log out of the computer before completing the dependent measures. If participants had written about an immoral experience, the cleansing manipulation decreased their amount of time spent volunteering to help a graduate student, presumably by washing away some of the residual guilt that would have motivated their compensatory prosocial behavior. But if participants had written about an emotionally neutral experience, the same cleansing manipulation *increased* their amount of time spent volunteering, presumably by conferring a sense of cleanliness that bore on the self, leading participants to see themselves as morally cleaner and to act in an identity-consistent manner.

These experiments highlight the importance of taking context into account when interpreting and predicting cleansing effects. More broadly, the situational conditions under which and the psychological mechanisms by which assimilation and contrast occur in cleansing effects will be important avenues for future research.

What Are the Limitations of the Present Meta-Analysis?

Limitations of Our Choice of Evidence Base

The present meta-analysis sought to provide a comprehensive quantitative assessment of causal effects regarding the psychological consequences and antecedents of physical cleansing. To that end, for a study to be included, it had to report (1) primary (2) quantitative data (3) involving human participants (4) in a true experiment that (5) either (a) manipulated physical cleansing and measured other psychological variables or (b) manipulated other psychological variables and measured outcomes directly about physical cleansing, (6) with an effect expected by the original researchers or the meta-analysts or both to be significant. The report had to (7) provide sufficient statistical information for effect size computation and (8) contain no ethical concern. Although these inclusion criteria served the focal goal of our meta-analysis, they inevitably meant the exclusion of certain other kinds of evidence that could be informative for the psychology of cleansing. For example, our evidence base did not include secondary data from panel studies, behavioral observations from qualitative research, ethological work on non-human primates (Spruijt et al., 1992), and correlational studies among human participants (e.g., linguistic patterns, cross-cultural comparisons, correlations between individual differences in cleansing and in other variables). By maintaining a tight focus on causal evidence afforded by human experiments, we accepted the cost of missing potentially relevant evidence from alternative approaches and populations.

Our literature search sought to identify experimental evidence reported in any year and in any language. The earliest year was 1989 and the latest 2023, a range of 35 years. This range is slightly misleading though because there was a gap of over a decade after the single report in 1989 and before the next report in 2002. The majority of reports were in 2010 and after (Figure S2). Using number of reports as a proxy for interest in the topic, there was a surge of interest in the mid 2010s, which subsided afterwards. Across all years, the vast majority of evidence we obtained was reported in English; only a few reports were in other languages (German and Chinese). As noted earlier (**Participant Demographics** under ***What Methodological Aspects of Cleansing Experiments Matter?***), a possible reason is that our database of choice (APA PsycInfo) and our other sources of studies may have over-emphasized research in the West and scholarly work written in English. Future research may address this limitation by focusing on non-Western databases and non-English reports. As far as the present meta-analytic dataset is concerned, the lopsided distribution of English vs. non-English reports rendered it impossible to rule out mono-language bias, though the availability of cleansing effects from different countries and regions may indirectly alleviate this concern to some extent.

Limitations of Our Approach to Evidence Synthesis

In the process of synthesizing effect sizes, we used a variety of methods to assess and correct for outliers and publication bias. It is worth reiterating that there is no single perfect method (hence our decision to use multiple methods and present their results comprehensively). Readers interested in the specific limitations of each method should consult prior methodological investigations (Carter et al., 2019; McShane et al., 2016; Stanley, 2017; van Aert et al., 2016).

Our moderator analyses were extensive in terms of the variety of moderators examined (theoretical, methodological, participant demographics, report characteristics). Because of the large number of moderators, however, to keep the presentation manageable, we limited our moderator analyses to three statistical approaches: RE meta-regressions, PET, and PEESE, all with robust variance estimates (RVE). RVE comes with various advantages (see ***Handling of Non-Independent Effect Sizes***), but also the disadvantage of limited statistical power in probing moderation, which means null effects in our moderator analyses need to be treated with caution (Coles et al., 2019). Certain moderators also had lopsided distribution of effect sizes such that there were many effect sizes at one level of the moderator but only a few effect sizes at another level of the moderator (Figure S7). In these cases, additional caution is warranted in interpreting significant vs. non-significant differences in effect sizes between different levels of the moderator. We encourage greater emphasis on effect size and lesser emphasis on statistical significance. This is especially true where statistical power is low, though also true in general.

Conclusions

Physical cleansing is a human universal. It runs through our day. It reduces chances of sickness and increases chances of survival, in normal times and especially during pandemics of contagious diseases. It also carries rich meanings that have piqued the interest of cognitive linguists, affective scientists, social and moral psychologists, cultural anthropologists, and religious scholars. Building on centuries of anecdotal observations and decades of correlational data, experimental research in the past 15 years has systematically investigated what psychological states trigger cleansing-related thoughts, feelings, and behaviors, and what psychological effects result from cleansing. The present meta-analysis takes stock of this entire body of cleansing effects to provide hitherto the most comprehensive evidence-based answers to a number of research questions.

First and foremost, how strong is the empirical foundation of cleansing effects? On the one hand, the synthesized effect size estimates were significant regardless of which of the eight methods of bias assessment and correction was used, even after identifying and removing outliers based on leave-one-out diagnostics and sensitivity analysis. On the other hand, in general the bias-corrected estimates were small in size, and the effects came from underpowered, non-preregistered studies (with a few exceptions). Both successful and unsuccessful replications exist, which often pertain to different original effects, rendering direct comparisons difficult. Still, it is worth pointing out that successful replications typically had smaller sample sizes than unsuccessful ones. Roughly half of the effects in unsuccessful replications came from preregistered studies; none of the effects in successful replications came from preregistered studies. These weaknesses should be sufficient reasons for future studies to follow contemporary norms of power analysis, preregistration, and replication. They should also serve a cautionary role in calibrating the observed patterns of cleansing effects and their theoretical implications.

Are cleansing effects stronger within the moral domain than beyond, as would be expected from the perspective of conceptual metaphors? No. Cleansing effects have been observed within the moral domain and beyond, with comparable effect sizes (Figure 4). The

domain-generality of cleansing effects is consistent with the perspective of cleansing as a grounded procedure of separation.

Within the moral domain, are cleansing effects stronger if sanctity/degradation concerns (which tend to be linked to disgust) are involved? It depends. Overall, cleansing effects exhibit similar effect sizes across different moral subdomains. However, psychological antecedents of cleansing do exhibit larger effect sizes for sanctity/degradation than for fairness/cheating, whereas psychological consequences of cleansing exhibit smaller effect sizes for sanctity/degradation than for fairness/cheating. That is, sanctity/degradation violations are particularly potent in triggering cleansing-related outcomes, whereas cleansing manipulations have relatively weak effects on sanctity/degradation-related outcomes. This nuanced pattern is consistent with several theoretical perspectives, including the role of disgust, domain-specific mapping between cleansing and morality, and cleansing as a grounded procedure of separation.

Beyond the moral domain, are cleansing effects stronger in some domains than in others? The data suggest that cleansing effects are observable regardless of whether the non-moral domains are social (i.e., involving others or society; e.g., gender, social presence, identity threat) or non-social (i.e., not involving others or society; e.g., postdecisional dissonance, luck, optimism), with comparable effect sizes.

Are cleansing effects unidirectional, or bidirectional? That is, does the literature show only psychological consequences of cleansing, as predicted by the perspective of metaphorical structuring or scaffolding? Or does it also show psychological antecedents of cleansing, as predicted by the perspectives of grounded cognition and neural reuse? The data support both psychological consequences and antecedents of cleansing, with similar effect sizes (Figure 8), suggesting that cleansing effects are bidirectional and symmetric in size overall.

What methodological aspects of cleansing experiments matter the most? Type of measure. Cleansing effects tend to be larger in size with behavioral measures than with mental and sensorimotor measures of psychological variables. Other than that, cleansing effects are observable and comparable in size across most of the coded methodological moderators, such as type of manipulation (actual experience, imagined or recalled experience, or conceptual activation), testing of mediation, testing of moderation, and inclusion of additional conditions in the focal factor.

Do cleansing effects vary by participant demographics and report characteristics? Yes. Cleansing effects show regional variation such that they are smaller in size, but much larger in count, among samples in the West than in the East or Middle East. Likewise, psychological consequences of cleansing are smaller in size, but far larger in count, among convenient samples (local, university-related, or online) than other or unspecified samples. Effect sizes are also smaller in non-peer-reviewed reports (relative to peer-reviewed ones) and in newer reports.

To distill all of our substantive meta-analytic findings into two sentences: On the whole, cleansing effects are small-to-medium in size, robust to various bias assessment and correction methods, especially strong on behavioral measures, weaker but still significant on other measures (judgment, feeling, thought process, and sensorimotor process). Cleansing effects, despite exhibiting heterogeneity between types of sample and report, are domain-general, bidirectional, and observable across diverse manipulations and study designs.

References

Note. References with an asterisk were part of the present meta-analysis.

American Psychological Association. (2020). *Meta-analysis reporting standards (MARS)*. https://apastyle.apa.org/jars/quant-table-9.pdf

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(04), 245–266. https://doi.org/10.1017/S0140525X10000853

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

*Allison-Godfrey, A., Bronson, K., Luby, E., Salazar, A., & Holmes, K. J. (2015, April 30). *Cleanliness reminders affect political attitudes.* http://psychfiledrawer.org/replication.php?attempt=MjIx

*Arbesfeld, J., Collins, T., Baldwin, D., & Daubman, K. (2014). *Clean Thoughts Lead to Less Severe Moral Judgment*. http://psychfiledrawer.org/replication.php?attempt=MTc3

Ashenburg, K. (2007). *The dirt on clean: An unsanitized history*. North Point Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99532147

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On Staying Grounded and Avoiding Quixotic Dead Ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

*Basáñez, T., Mirshafiei, E., & Farías-Yapur, A. (2019). *Effect of imagining a suspicious smell on mood, trust, and scrutiny of information*.

*Bastian, B., Bain, P., Buhrmester, M. D., Gomez, A., Vazquez, A., Knight, C. G., & Swann, W. B. (2015). Moral Vitalism: Seeing Good and Evil as Real, Agentic Forces. *Personality and Social Psychology Bulletin*, *41*(8), 1069–1081. https://doi.org/10.1177/0146167215589819

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Becker, B. J. (2005). Failsafe *N* or File-Drawer Number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis* (1st ed., pp. 111–125). Wiley. https://doi.org/10.1002/0470870168.ch7

*Besman, M., Dubensky, C., Dunsmore, L., & Daubman, K. (2013). *Cleanliness Primes Less Severe Moral Judgments*. http://psychfiledrawer.org/replication.php?attempt=MTQ5

*Beyfus, T. A., Dawson, N. L., Danner, C. H., Rawal, B., Gruber, P. E., & Petrou, S. P. (2016). The use of passive visual stimuli to enhance compliance with handwashing in a perioperative setting. *American Journal of Infection Control*, *44*(5), 496–499. https://doi.org/10.1016/j.ajic.2015.12.021

*Bilekli, I., & Inozu, M. (2018). Mental contamination: The effects of religiosity. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 43–50. https://doi.org/10.1016/j.jbtep.2017.08.001

*Bilz, K. (2012). Dirty Hands or Deterrence? An Experimental Examination of the Exclusionary Rule: The Exclusionary Rule in Law. *Journal of Empirical Legal Studies*, *9*(1), 149–171. https://doi.org/10.1111/j.1740-1461.2011.01250.x

Blackman, A. M. (1918). Some Notes on the Ancient Egyptian Practice of Washing the Dead. *The Journal of Egyptian Archaeology*, *5*(2), 117. https://doi.org/10.2307/3853730

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology*, *42*, 319–373.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2* [Computer software]. Biostat. https://www.meta-analysis.com/

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Borman, G. D., & Grigg, J. A. (2009). Visual and narrative interpretation. In *The handbook of research synthesis and meta-analysis, 2nd ed* (pp. 497–519). Russell Sage Foundation.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*(1), 1–28. https://doi.org/10.1016/S0010-0277(99)00073-6

Boyce, J. M., & Pittet, D. (2002). Guideline for hand hygiene in health-care settings: Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Infection Control & Hospital Epidemiology*, *23*(S12), S3–S40. https://doi.org/10.1086/503164

*Briñol, P., Gascó, M., Petty, R. E., & Horcajo, J. (2013). Treating Thoughts as Material Objects Can Increase or Decrease Their Impact on Evaluation. *Psychological Science*, *24*(1), 41–47. https://doi.org/10.1177/0956797612449176

*Burnham, B. R. (2020). Are liberals really dirty? Two failures to replicate Helzer and Pizarro's (2011) study 1, with meta-analysis. *Journal of Personality and Social Psychology*, *119*(6), e38–e42. https://doi.org/10.1037/pspa0000238

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of psychophysiology*. Cambridge University Press. http://www.dawsonera.com/depp/reader/protected/external/AbstractView/S9780511320736

*Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A Constructionist Review of Morality and Emotions: No Evidence for Specific Links Between Moral Content and Discrete Emotions. *Personality and Social Psychology Review*, *19*(4), 371–394. https://doi.org/10.1177/1088868314566683

*Cardinale Lagomarsino, B., Gutman, M., Freira, L., Lanzalot, M. L., Lauletta, M., Malchik, L. E., Montaño Campos, F., Pacini, B., Rossi, M. A., & Valencia, C. (2017). Peer pressure: Experimental evidence from restroom behavior. *Economic Inquiry*, *55*(3), 1579–1584. https://doi.org/10.1111/ecin.12437

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

*Chan, E. Y. (2019). Dirty weekends and personal hygiene products: The embodiment of casual sex in marketing. *Psychology & Marketing*. https://doi.org/10.1002/mar.21198

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300–327. https://doi.org/10.1037/a0030964

*Cho, H. (2019). Brand name fluency and perceptions of water purity and taste. *Food Quality and Preference*, *71*, 21–24. https://doi.org/10.1016/j.foodqual.2018.05.002

*Ciccarelli, T., Soberman, M., Leshuk, T., Cole, H., Afreen, F., & Manwell, L. A. (2021). Is cleanliness next to abstinence? The effect of cleanliness priming on attitudes towards

harm reduction strategies for people with substance use disorders. I*nternational Journal of Psychology, 56*(2), 322–330. https://doi.org/10.1002/ijop.12703
Coburn, K. M., & Vevea, J. L. (2019). *weightr: Estimating Weight-Function Models for Publication Bias* (2.0.2) [Computer software]. https://CRAN.R-project.org/package=weightr
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441
Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin*, *145*(6), 610–651. https://doi.org/10.1037/bul0000194
*Coughtrey, A. E., Shafran, R., & Rachman, S. J. (2014). The spontaneous decay and persistence of mental contamination: An experimental analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 90–96. https://doi.org/10.1016/j.jbtep.2013.09.001
*Cougle, J. R., Goetz, A. R., Hawkins, K. A., & Fitch, K. E. (2012). Guilt and Compulsive Washing: Experimental Tests of Interrelationships. *Cognitive Therapy and Research*, *36*(4), 358–366. https://doi.org/10.1007/s10608-011-9359-x
*Cramwinckel, F. M., De Cremer, D., & van Dijke, M. (2013). Dirty Hands Make Dirty Leaders?! The Effects of Touching Dirty Objects on Rewarding Unethical Subordinates as a Function of a Leader's Self-Interest. *Journal of Business Ethics*, *115*(1), 93–100. https://doi.org/10.1007/s10551-012-1385-4
*Cramwinckel, F. M., van Dijk, E., Scheepers, D., & van den Bos, K. (2013). The threat of moral refusers for one's self-concept and the protective function of physical cleansing. *Journal of Experimental Social Psychology*, *49*(6), 1049–1058. https://doi.org/10.1016/j.jesp.2013.07.009
*Cui, Y. (Gina), Errmann, A., Kim, J., Seo, Y., Xu, Y., & Zhao, F. (2020). Moral Effects of Physical Cleansing and Pro-environmental Hotel Choices. *Journal of Travel Research*, *59*(6), 1105–1118. https://doi.org/10.1177/0047287519872821
*de Groot, J., Abrahamse, W., & Vincent, S. (2013). Thou Shalt Not Steal: Effects of Normative Cues on Attitudes towards Theft. *Psychology*, *04*(04), 438–444. https://doi.org/10.4236/psych.2013.44062
*De Los Reyes, A., Aldao, A., Kundey, S. M. A., Lee, B. G., & Molina, S. (2012). Compromised decision making and the effects of manipulating physical states on human judgments. *Journal of Clinical Psychology*, *68*(1), 1–7. https://doi.org/10.1002/jclp.20851
Del Re, A. C., & Hoyt, W. (2022). *MAd: Meta-Analysis with Mean Differences* (0.8-3) [Computer software]. https://CRAN.R-project.org/package=MAd
*Denison, S., Lee, A., Lin, J., & Prince, A. (2018). *The Macbeth effect and motor modality: An attempted replication* [Poster]. Department Poster Session, Swarthmore College.
*Denke, C., Rotte, M., Heinze, H.-J., & Schaefer, M. (2014). Lying and the Subsequent Desire for Toothpaste: Activity in the Somatosensory Cortex Predicts Embodiment of the Moral-Purity Metaphor. *Cerebral Cortex*, *26*(2), 477–484. https://doi.org/10.1093/cercor/bhu170
*D'Olimpio, F., & Mancini, F. (2014). Role of Deontological Guilt in Obsessive-Compulsive Disorder–Like Checking and Washing Behaviors. *Clinical Psychological Science*, *2*(6), 727–739. https://doi.org/10.1177/2167702614529549
Dong, P., & Lee, S. W. S. (2017). Embodiment as procedures: Physical cleansing changes goal priming effects. *Journal of Experimental Psychology: General*, *146*(4), 592–605. https://doi.org/10.1037/xge0000285
*Doron, G., Sar-El, D., & Mikulincer, M. (2012). Threats to moral self-perceptions trigger obsessive compulsive contamination-related behavioral tendencies. *Journal of Behavior*

Therapy and Experimental Psychiatry, *43*(3), 884–890. https://doi.org/10.1016/j.jbtep.2012.01.002
Douglas, M. (1966). *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*. Routledge.
Duval, S., & Tweedie, R. (2000a). A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905
Duval, S., & Tweedie, R. (2000b). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x
*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*(1), 91–98. https://doi.org/10.1080/01973533.2013.856792
*Eckmanns, T., Bessert, J., Behnke, M., Gastmeier, P., & Rüden, H. (2006). Compliance With Antiseptic Hand Rub Use in Intensive Care Units The Hawthorne Effect. *Infection Control & Hospital Epidemiology*, *27*(9), 931–934. https://doi.org/10.1086/507294
*Edwards, D., Monk-Turner, E., Poorman, S., Rushing, M., Warren, S., & Willie, J. (2002). Predictors of hand-washing behavior. *Social Behavior and Personality: An International Journal*, *30*(8), 751–756. https://doi.org/10.2224/sbp.2002.30.8.751
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629
Eliade, M. (1996). *Patterns in Comparative Religion*. University of Nebraska Press. (Original work published 1958)
*Elliott, C. M., & Radomsky, A. S. (2009). Analyses of mental contamination: Part I, experimental manipulations of morality. *Behaviour Research and Therapy*, *47*(12), 995–1003. https://doi.org/10.1016/j.brat.2009.03.004
*Elliott, C. M., & Radomsky, A. S. (2012). Mental contamination: The effects of imagined physical dirt and immoral behaviour. *Behaviour Research and Therapy*, *50*(6), 422–427. https://doi.org/10.1016/j.brat.2012.03.007
*Fairbrother, N., Newth, S. J., & Rachman, S. (2005). Mental pollution: Feelings of dirtiness without physical contact. *Behaviour Research and Therapy*, *43*(1), 121–130. https://doi.org/10.1016/j.brat.2003.12.005
*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*(2), 21–29.
Feder, Y. (2016). Contamination Appraisals, Pollution Beliefs, and the Role of Cultural Inheritance in Shaping Disease Avoidance Behavior. *Cognitive Science*, *40*(6), 1561–1585. https://doi.org/10.1111/cogs.12293
Fisher, Z., Tipton, E., & Zhipeng, H. (2023). *robumeta: Robust Variance Meta-Regression* (2.1) [Computer software]. https://cran.r-project.org/web/packages/robumeta/index.html
Fitouchi, L., André, J.-B., Baumard, N., & Nettle, D. (2022). *Harmless bodily pleasures are moralized because they are perceived as reducing self-control and cooperativeness* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fzv43
*Florack, A., Kleber, J., Busch, R., & Stöhr, D. (2014). Detaching the ties of ownership: The effects of hand washing on the exchange of endowed products. *Journal of Consumer Psychology*, *24*(2), 284–289. https://doi.org/10.1016/j.jcps.2013.09.010
Galoni, C., & Noseworthy, T. J. (2015). Does dirty money influence product valuations? *Journal of Consumer Psychology*, *25*(2), 304–310. https://doi.org/10.1016/j.jcps.2014.11.002

*Gámez, E., Díaz, J. M., & Marrero, H. (2011). The Uncertain Universality of the Macbeth Effect with a Spanish Sample. *The Spanish Journal of Psychology*, *14*(01), 156–162. https://doi.org/10.5209/rev_SJOP.2011.v14.n1.13

Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

*Gilchrist, P. T., & Schnall, S. (2018). The paradox of moral cleansing: When physical cleansing leads to increased contamination concerns. *Journal of Behavior Therapy and Experimental Psychiatry*, *61*, 38–44. https://doi.org/10.1016/j.jbtep.2018.06.002

Giner-Sorolla, R. (2018, January 24). Powering Your Interaction. *Approaching Significance*. https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

Giner-Sorolla, R., & Sabo, J. S. (2016). Disgust in the moral realm: Do all roads lead to character? In R. Duschinsky, S. Schnall, & D. H. Weiss (Eds.), *Purity and Danger Now* (pp. 87–102). Routledge.

*Golec de Zavala, A., Waldzus, S., & Cypryanska, M. (2014). Prejudice towards gay men and a need for physical cleansing. *Journal of Experimental Social Psychology*, *54*, 1–10. https://doi.org/10.1016/j.jesp.2014.04.001

*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*(6), 1356–1360. https://doi.org/10.1016/j.jesp.2012.07.001

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Elsevier. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

*Grant, A. M., & Hofmann, D. A. (2011). It's Not All About Me: Motivating Hand Hygiene Among Health Care Professionals by Focusing on Patients. *Psychological Science*, *22*(12), 1494–1499. https://doi.org/10.1177/0956797611419172

Gray, K., Waytz, A., & Young, L. (2012). The Moral Dyad: A Fundamental Template Unifying Moral Judgment. *Psychological Inquiry*, *23*(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In *The handbook of research synthesis and meta-analysis, 2nd ed* (pp. 417–433). Russell Sage Foundation.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319–359). Academic.

*Haggard, M. C., Rowatt, W., & Yelderman, L. (2014, February 15). *Cleanliness is Like Godliness: Harsher Judgments of Out-groups Follow Cleansing after Disgust* [Poster]. Annual meeting of the Society for Personality and Social Psychology, Austin, TX.

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203–217. https://doi.org/10.1037/1082-989X.6.3.203

Hedges, L. V., & Pigott, T. D. (2004). The Power of Statistical Tests for Moderators in Meta-Analysis. *Psychological Methods*, *9*(4), 426–445. https://doi.org/10.1037/1082-989X.9.4.426

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hedges, L. V., & Vevea, J. L. (1996). Estimating Effect Size Under Publication Bias: Small Sample Properties and Robustness of a Random Effects Selection Model. *Journal of Educational and Behavioral Statistics*, *21*(4), 299–332. https://doi.org/10.3102/10769986021004299

*Helzer, E. G., & Pizarro, D. A. (2011). Dirty Liberals! Reminders of Physical Cleanliness Influence Moral and Political Attitudes. *Psychological Science*, *22*(4), 517–522. https://doi.org/10.1177/0956797611402514

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

*Herba, J. K., & Rachman, S. (2007). Vulnerability to mental contamination. *Behaviour Research and Therapy*, *45*(11), 2804–2812. https://doi.org/10.1016/j.brat.2007.07.010

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, *19*(4), 323–340. https://doi.org/10.1016/0022-1031(83)90026-4

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and Self-Control: How They Are Intertwined and Where They Differ. *Current Directions in Psychological Science*, *27*(4), 286–291. https://doi.org/10.1177/0963721418759317

Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells Like Clean Spirit: Nonconscious Effects of Scent on Cognition and Behavior. *Psychological Science*, *16*(9), 689–693. https://doi.org/10.1111/j.1467-9280.2005.01597.x

Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as Moral Amplifiers: An Appraisal Tendency Approach to the Influences of Distinct Emotions upon Moral Judgment. *Emotion Review*, *3*(3), 237–244. https://doi.org/10.1177/1754073911402384

Hoy, S. M. (1995). *Chasing dirt the American pursuit of cleanliness*. Oxford University Press. http://www.ebrary.com

*Huang, J. L. (2014). Does cleanliness influence moral judgments? Response effort moderates the effect of cleanliness priming on moral judgments. *Frontiers in Psychology*, *5*, 1276. https://doi.org/10.3389/fpsyg.2014.01276

*Huang, J., Ackerman, J., & Sedlovskaya, A. (2012). *Protection from Disease Attenuates Contamination-Related Biases*. Society for Consumer Psychology 2012 Annual Winter Conference, 162–163.

*Huang, J. Y., Ackerman, J. M., & Sedlovskaya, A. (2017). (De)contaminating product preferences: A multi-method investigation into pathogen threat's influence on used product preferences. *Journal of Experimental Social Psychology, 70*, 143–152. https://doi.org/10.1016/j.jesp.2017.01.001

*Huang, J. Y., Sedlovskaya, A., Ackerman, J. M., & Bargh, J. A. (2011). Immunizing Against Prejudice: Effects of Disease Protection on Attitudes Toward Out-Groups. *Psychological Science*, *22*(12), 1550–1556. https://doi.org/10.1177/0956797611417261

*Huangfu, G., Lv, F., Sheng, C., & Shi, X. (2017). Effect of workplace environment cleanliness on judgment of counterproductive work behavior. *Social Behavior and Personality: An International Journal*, *45*(4), 599–604. https://doi.org/10.2224/sbp.6083

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*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Hunter, J. E., & Schmidt, F. L. (2000). Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge. *International Journal of Selection and Assessment*, *8*(4), 275–292. https://doi.org/10.1111/1468-2389.00156

IJzerman, H., & Koole, S. L. (2011). From perceptual rags to metaphoric riches—Bodily, social, and cultural constraints on sociocognitive metaphors: Comment on Landau, Meier, and Keefer (2010). *Psychological Bulletin*, *137*(2), 355–361. https://doi.org/10.1037/a0022373

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*(605), F236–F265. https://doi.org/10.1111/ecoj.12461

Iyengar, S., & Greenhouse, J. B. (1988). Selection Models and the File Drawer Problem. *Statistical Science*, *3*(1), 109–117.

*Jia, M. (Liam), Li, X., & Krishna, A. (2018). "But, will you think it's important to use mouthwash?" How visual communication of a set impacts perceived set completeness and item importance. *Advances in Consumer Research*, *46*, 154–159. https://www.acrwebsite.org/volumes/v46/acr_vol46_2412321.pdf

*Jiang, Y., & Gao, L. (2015). *I am feeling the urge to clean out my closet: How the mere thought of change influences product disposal decisions*.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin*, *147*(1), 1–15. https://doi.org/10.1037/bul0000317

Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine*, *233*, 237–251. https://doi.org/10.1016/j.socscimed.2019.05.035

*Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014a). Does Cleanliness Influence Moral Judgments?: A Direct Replication of Schnall, Benton, and Harvey (2008). *Social Psychology*, *45*(3), 209–215. https://doi.org/10.1027/1864-9335/a000186

*Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014b). *Cleanliness primes do not influence moral judgment*. http://psychfiledrawer.org/replication.php?attempt=MTcy

*Kalanthroff, E., Aslan, C., & Dar, R. (2015). Washing away your sins will set your mind free: Physical cleansing modulates the effect of threatened morality on executive control. *Cognition and Emotion*, 1–8. https://doi.org/10.1080/02699931.2015.1086313

Kampf, G., & Kramer, A. (2004). Epidemiologic background of hand hygiene and evaluation of the most important agents for scrubs and rubs. *Clinical Microbiology Reviews*, *17*(4), 863–893. https://doi.org/10.1128/CMR.17.4.863-893.2004

*Kaspar, K. (2012). Washing one's hands after failure enhances optimism but hampers future performance. *Social Psychological and Personality Science*, *4*(1), 69–73. https://doi.org/10.1177/1948550612443267

*Kaspar, K., & Cames, S. (2016). Cognitions about bodily purity attenuate stress perception. *Scientific Reports*, *6*, 38829. https://doi.org/10.1038/srep38829

*Kaspar, K., & Jahn, C. (2016). Physical cleansing biases recognition performance for (im)moral social issues. *Studia Psychologica*, *58*(4), 307–321. https://doi.org/10.21909/sp.2016.04.725

*Kaspar, K., & Klane, A. (2016). Embodied information affects judgements about politicians: The impact of haptic weight sensations and hand cleansing. *International Journal of Psychology*. https://doi.org/10.1002/ijop.12366

*Kaspar, K., Krapp, V., & König, P. (2015). Hand Washing Induces a Clean Slate Effect in Moral Judgments: A Pupillometry and Eye-Tracking Study. *Scientific Reports*, *5*, 10471. https://doi.org/10.1038/srep10471

*Kaspar, K., & Teschlade, L. (2016). Does Physical Purity License Moral Transgressions or Does it Increase the Tendency towards Moral Behavior? *Current Psychology*. https://doi.org/10.1007/s12144-016-9483-6

*Khan, M., & Grisham, J. R. (2018). Wiping your conscience clean: Investigating the Macbeth effect in individuals with high obsessive-compulsive contamination concerns. *Journal of Experimental Psychopathology*, *9*(3), 1–10. https://doi.org/10.1177/2043808718786595

*Kim, C., & Huh, Y. E. (2019). Expiration date perception and food disposal decision. *Advances in Consumer Research*, *47*, 158–163. https://www.acrwebsite.org/volumes/v47/acr_vol47_2551092.pdf

*Kim, J., Kim, J.-E., & Park, J. (2018). Effects of physical cleansing on subsequent unhealthy eating. *Marketing Letters*. https://doi.org/10.1007/s11002-018-9458-5

*Kim, T. W., Duhachek, A., Briñol, P., Lee, S. W. S., & Petty, R. E. (2019). *The Meaning of Physical Actions Moderate the Impact of Cleansing on Guilt*.

*Kim, Y.-H., & Cohen, D. (2010). Information, Perspective, and Judgments About the Self in Face and Dignity Cultures. *Personality and Social Psychology Bulletin*, *36*(4), 537–550. https://doi.org/10.1177/0146167210362398

*King, D., Vlaev, I., Everett-Thomas, R., Fitzpatrick, M., Darzi, A., & Birnbach, D. J. (2016). "Priming" hand hygiene compliance in clinical environments. *Health Psychology*, *35*(1), 96–101. https://doi.org/10.1037/hea0000239

*Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

*Körner, A., & Strack, F. (2019). Conditions for the clean slate effect after success or failure. *The Journal of Social Psychology*, *159*(1), 92–105. https://doi.org/10.1080/00224545.2018.1454881

*Kwok, P. L. A. (2010). *Mental Contamination: A Replication and Extension of the "Dirty Kiss" Experiment* [Ph.D., The Chinese University of Hong Kong]. http://search.proquest.com/psycinfo/docview/902185476/abstract/9217F407B1514B61PQ/1

Lakens, D. (2020, March 29). The 20% Statistician: Effect Sizes and Power for Interactions in ANOVA Designs. *The 20% Statistician*. http://daniellakens.blogspot.com/2020/03/effect-sizes-and-power-for-interactions.html

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Landau, M. J., Keefer, L. A., & Meier, B. P. (2011). Wringing the perceptual rags: Reply to IJzerman and Koole (2011). *Psychological Bulletin*, *137*(2), 362–365. https://doi.org/10.1037/a0022457

Landy, J. F., & Goodwin, G. P. (2015). Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence. *Perspectives on Psychological Science*, *10*(4), 518–536. https://doi.org/10.1177/1745691615583128

*Lapinski, M. K., Maloney, E. K., Braz, M., & Shulman, H. C. (2013). Testing the Effects of Social Norms and Behavioral Privacy on Hand Washing: A Field Experiment. *Human Communication Research*, *39*(1), 21–46. https://doi.org/10.1111/j.1468-2958.2012.01441.x

*Lee, A., & Ji, L.-J. (2015, February). *Physical cleansing moderates the link between investment of labor and feelings of ownership* [Poster]. Annual meeting of the Society for Personality and Social Psychology, Long Beach, CA.

Lee, S. W. S., Chen, K., Ma, C., & Hoang, J. (2023, September 28). Wipe It Off: A Meta-Analytic Review of the Psychological Consequences and Antecedents of Physical Cleansing. Retrieved from osf.io/xv2wj

Lee, S. W. S., & Ellsworth, P. C. (2013). Maggots and morals: Physical disgust is to fear as moral disgust is to anger. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 271–280). Oxford University Press.

*Lee, S. W. S., Millet, K., Grinstein, A., Pauwels, K. H., Johnston, P. R., Volkov, A. E., & van der Wal, A. J. (2023). Actual Cleaning and Simulated Cleaning Attenuate Psychological and Physiological Effects of Stressful Events. *Social Psychological and Personality Science*, *14*(4), 381–394. https://doi.org/10.1177/19485506221099428

*Lee, S. W. S., & Schwarz, N. (2010a). Dirty hands and dirty mouths: Embodiment of the moral-purity metaphor is specific to the motor modality involved in moral transgression. *Psychological Science*, *21*(10), 1423–1425. https://doi.org/10.1177/0956797610382788

*Lee, S. W. S., & Schwarz, N. (2010b). Washing away postdecisional dissonance. *Science*, *328*(5979), 709. https://doi.org/10.1126/science.1186799

Lee, S. W. S., & Schwarz, N. (2011). Wiping the slate clean: Psychological consequences of physical cleansing. *Current Directions in Psychological Science*, *20*(5), 307–311. https://doi.org/10.1177/0963721411422694

Lee, S. W. S., & 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*(5), 737–749. https://doi.org/10.1037/a0029708

Lee, S. W. S., & Schwarz, N. (2016). Clean-moral effects and clean-slate effects: Physical cleansing as an embodied procedure of psychological separation. In R. Duschinsky, S. Schnall, & D. H. Weiss (Eds.), *Purity and Danger Now: New Perspectives* (pp. 136–161). Routledge.

Lee, S. W. S., & Schwarz, N. (2021). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions [Target Article]. *Behavioral and Brain Sciences*, 1–78. https://doi.org/10.1017/S0140525X20000308

*Lee, S. W. S., Tang, H., Wan, J., Mai, X., & Liu, C. (2015). A cultural look at moral purity: Wiping the face clean. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00577

Leung, L.-W. S., & Borst, J. G. G. (1987). Electrical activity of the cingulate cortex. I. Generating mechanisms and relations to behavior. *Brain Research*, *407*(1), 68–80. https://doi.org/10.1016/0006-8993(87)91220-0

*Li, C., Liu, L., Zheng, W., Dang, J., & Liang, Y. (2017). A clean self reduces bribery intent. *International Journal of Psychology*. https://doi.org/10.1002/ijop.12452

Life.Church. (n.d.). *Soul Detox*. Retrieved February 15, 2019, from https://www.bible.com/reading-plans/257-soul-detox

*Liljenquist, K., Zhong, C.-B., & Galinsky, A. D. (2010). The smell of virtue: Clean scents promote reciprocity and charity. *Psychological Science*, *21*(3), 381–383. https://doi.org/10.1177/0956797610361426

*Lobel, T. E., Cohen, A., Kalay Shahin, L., Malov, S., Golan, Y., & Busnach, S. (2015). Being Clean and Acting Dirty: The Paradoxical Effect of Self-Cleansing. *Ethics & Behavior*, *25*(4), 307–313. https://doi.org/10.1080/10508422.2014.931230

*Ma, C., & Lee, S. W. S. (2022). *Does physical cleansing change goal priming effects? Preregistered replications*.

*Marotta, M., & Bohner, G. (2013, September). *Dissonanz abwaschen, Dissonanz reinreiben: Symbolische Abschwächung vs. Verstärkung von Dissonanz nach Entscheidungen [Washing dissonance off, rubbing dissonance in: Symbolic attenuation and intensification of postdecisional dissonance]*. Tagung der Fachgruppe Sozialpsychologie, Hagen, Germany.

*Martins, C. M., Block, L. G., & Dahl, D. W. (2015). Can Hand Washing Influence Hedonic Food Consumption? *Psychology & Marketing*, *32*(7), 742–750. https://doi.org/10.1002/mar.20814

Mason, J. R., Reidinger, Jr., R. F., & Katz, Y. (1985). Flavor avoidance expressed in grooming by pine voles (Microtus pinetorum): Importance of context and hormonal factors. *Physiology & Behavior*, *35*(6), 979–983. https://doi.org/10.1016/0031-9384(85)90269-0

*Maury, E., Moussa, N., Lakermi, C., Barbut, F., & Offenstadt, G. (2006). Compliance of health care workers to hand hygiene: Awareness of being observed is important. *Intensive Care Medicine*, *32*(12), 2088–2089. https://doi.org/10.1007/s00134-006-0398-9

*McGraw, P. A., Schwartz, J. A., & Tetlock, P. E. (2012). From the Commercial to the Communal: Reframing Taboo Trade-offs in Religious and Pharmaceutical Marketing. *Journal of Consumer Research, 39*(1), 157–173. https://doi.org/10.1086/662070

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.009

*Meng, M. D., & Gamlin, J. (2023). Sexually explicit advertisements boost consumer recycling due to moral cleansing goal activation. *Marketing Letters*. https://doi.org/10.1007/s11002-023-09689-0

*Menzies, R. E., & Dar-Nimrod, I. (2017). Death anxiety and its relationship with obsessive-compulsive disorder. *Journal of Abnormal Psychology*, *126*(4), 367–377. https://doi.org/10.1037/abn0000263

Michael, A. (1979). *Radhakrishnan on Hindu Moral Life and Action*. Concept.

*Millar, J. F. A., Salkovskis, P. M., & Brown, C. (2016). Mental contamination in the "dirty kiss": Imaginal betrayal or bodily fluids? *Journal of Obsessive-Compulsive and Related Disorders, 8*, 70–74. https://doi.org/10.1016/j.jocrd.2015.12.004

*Mishra, N., & Whitley, S. (2020). *"Who" You Rent From Matters: The Differential Effects in Perceived Ownership When Renting*. ACR North American Advances, NA-48. https://www.acrwebsite.org/volumes/2662044/volumes/v48/NA-48

*Mitakakis, C. (2012). *Cleansing effects in consumer behavior* [Ph.D., City University of New York]. http://search.proquest.com/psycinfo/docview/1018099111/abstract/4C68D191FFB4DE7PQ/1

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Mooijman, M., Meindl, P., Oyserman, D., Monterosso, J., Dehghani, M., Doris, J. M., & Graham, J. (2018). Resisting temptation for the good of the group: Binding moral values and the moralization of self-control. *Journal of Personality and Social Psychology*, *115*(3), 585–599. https://doi.org/10.1037/pspp0000149

*Moris, I., & Melzer, A. (2019). Replication of Gollwitzer & Melzer, 2012 [Unpublished undergraduate thesis]. University of Luxembourg.
*Morrin, M., Lwin, M. O., Pham, N., & Rank-Christman, T. (2014). Ethical mindsets and moral transgression classification: When cleansing licenses consumer indulgence. *Advances in Consumer Research*, *42*, 807. https://www.acrwebsite.org/volumes/v42/acr_v42_18005.pdf
*Morrin, M., & Rank-Christman, T. (2013). Losing Control: When Physical Cleansing Intensifies Cheating for High Self-Controls. *Advances in Consumer Research*, *41*, 804. http://www.acrwebsite.org/volumes/1014736/volumes/v41/NA-41
Morrison, D. (2011, November 23). *India Stems Tide of Pollution Into Ganges River*. National Geographic News. https://news.nationalgeographic.com/news/2011/11/111123-india-ganges-river-pollution/
*Moscatiello, T., & Nagel, E. (2014, May 2). *The effects of cleansing on risk taking* [Poster]. Celebration, Gettysburg, PA. http://cupola.gettysburg.edu/cgi/viewcontent.cgi?article=1592&context=celebration
*Munger, K., & Harris, S. J. (1989). Effects of an Observer on Handwashing in a Public Restroom. *Perceptual and Motor Skills*, *69*(3–1), 733–734. https://doi.org/10.1177/003151258906 93-104
Mussweiler, T., Rüter, K., & Epstude, K. (2004). The Ups and Downs of Social Comparison: Mechanisms of Assimilation and Contrast. *Journal of Personality and Social Psychology*, *87*(6), 832–844. https://doi.org/10.1037/0022-3514.87.6.832
National Institute of Mental Health. (n.d.). *Obsessive-Compulsive Disorder*. Retrieved March 7, 2016, from http://www.nimh.nih.gov/health/topics/obsessive-compulsive-disorder-ocd/index.shtml
Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology*, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836
Nemeroff, C., & Rozin, P. (1994). The Contagion Concept in Adult Thinking in the United States: Transmission of Germs and of Interpersonal Influence. *Ethos*, *22*(2), 158–186. https://doi.org/10.1525/eth.1994.22.2.02a00020
Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157
*Nussinson, R., Amit, A., & Weinstein, Z. (2019). *Metaphorically representing distinctiveness threat as a threat of physical contamination*.
Olivera La Rosa, A., & Rosselló Mir, J. (2013). On the relationships between disgust and morality: A critical review. *Psicothema*, *25*(2), 222–226. https://doi.org/10.7334/psicothema2012.159
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716
Orwin, R. G. (1983). A Fail-Safe *N* for Effect Size in Meta-Analysis. *Journal of Educational Statistics*, *8*(2), 157–159. https://doi.org/10.3102/10769986008002157
*Ottaviani, C., Collazzoni, A., D'Olimpio, F., Moretta, T., & Mancini, F. (2018). I obsessively clean because deontological guilt makes me feel physiologically disgusted! *Journal of Obsessive-Compulsive and Related Disorders*. https://doi.org/10.1016/j.jocrd.2018.01.004
*Park, J., & Ishii, H. (2013). *Do Not Wash Your Hands When You Eat Junk Food: The Possibility of a Reverse Macbeth Effect in Consumer Behavior*. Advances in Consumer Research, 41, 805.

*Pfattheicher, S., Strauch, C., Diefenbacher, S., & Schnuerch, R. (2018). A field study on watching eyes and hand hygiene compliance in a public restroom. *Journal of Applied Social Psychology*, *48*(4), 188–194. https://doi.org/10.1111/jasp.12501

*Pilotti, M. A. E., & El Alaoui, K. F. (2018). A study of memory, evaluation, and choice with an (un)clean conscience. *Australian Journal of Psychology*. https://doi.org/10.1111/ajpy.12223

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, *66*, 29–38. https://doi.org/10.1016/j.jesp.2015.09.012

Pittet, D., Allegranzi, B., Boyce, J., & World Health Organization World Alliance for Patient Safety First Global Patient Safety Challenge Core Group of Experts. (2009). The World Health Organization guidelines on hand hygiene in health care and their consensus recommendations. *Infection Control & Hospital Epidemiology*, *30*(07), 611–622. https://doi.org/10.1086/600379

Pizarro, D., Inbar, Y., & Helion, C. (2011). On Disgust and Moral Judgment. *Emotion Review*, *3*(3), 267–268. https://doi.org/10.1177/1754073911402394

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and Reproducibility of Meta-Analyses in Psychology: A Meta-Review. *Perspectives on Psychological Science*, *15*(4), 1026–1041. https://doi.org/10.1177/1745691620906416

Pole, N. (2013). Disgust discussed: Introduction to the special section. *Psychological Bulletin*, *139*(2), 269–270. https://doi.org/10.1037/a0032073

*Poon, K.-T. (2019). Do You Reap What You Sow? The Effect of Cyberostracism on Moral Impurity. *Basic and Applied Social Psychology*, *41*(2), 132–146. https://doi.org/10.1080/01973533.2019.1585353

*Preston, J. L., & Ritter, R. S. (2012). Cleanliness and godliness: Mutual association between two kinds of personal purity. *Journal of Experimental Social Psychology*, *48*(6), 1365–1368. https://doi.org/10.1016/j.jesp.2012.05.015

*Prokosch, M. L., Gassen, J., Ackerman, J. M., & Hill, S. E. (2019). Caution in the time of cholera: Pathogen threats decrease risk tolerance. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000160

Pustejovsky, J. (2022). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (0.5.8) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

R Core Team. (2023). *R: A language and environment for statistical computing* (4.3.1) [R; R Foundation for Statistical Computing]. https://www.R-project.org/

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In *The handbook of research synthesis and meta-analysis, 2nd ed* (pp. 295–315). Russell Sage Foundation.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, *103*(1), 111–120. https://doi.org/10.1037/0033-2909.103.1.111

*Reicher, S. D., Templeton, A., Neville, F., Ferrari, L., & Drury, J. (2016). Core disgust is attenuated by ingroup relations. *Proceedings of the National Academy of Sciences*, *113*(10), 2631–2635. https://doi.org/10.1073/pnas.1517027113

Reidinger, R. F., Jr., Beauchamp, G. K., & Barth, M. (1982). Conditioned aversion to a taste perceived while grooming. *Physiology & Behavior*, *28*(4), 715–723. https://doi.org/10.1016/0031-9384(82)90057-9

*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*(2), 224–229. https://doi.org/10.1177/2167702613485565

Ritter, R. S., & Preston, J. L. (2011). Gross gods and icky atheism: Disgust responses to rejected religious beliefs. *Journal of Experimental Social Psychology*, *47*(6), 1225–1230. https://doi.org/10.1016/j.jesp.2011.05.006

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468. https://doi.org/10.1111/j.0014-3820.2005.tb01004.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *99*(3), 400–406. https://doi.org/10.1037/0033-2909.99.3.400

*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*(5), 722–741. https://doi.org/10.1007/s11031-015-9487-9

Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review*, *94*(1), 23–41. https://doi.org/10.1037/0033-295X.94.1.23

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 757–776). Guilford.

Russell, P. S., & Giner-Sorolla, R. (2013). Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological Bulletin*, *139*(2), 328–351. https://doi.org/10.1037/a0029319

*Sassenrath, C., Diefenbacher, S., Siegel, A., & Keller, J. (2015). A person-oriented approach to hand hygiene behavior: Emotional Empathy Fosters Hand Hygiene Practice. *Psychology & Health*, 1–43. https://doi.org/10.1080/08870446.2015.1088945

*Schaefer, M. (2017). Morality and soap in engineers and social scientists: The Macbeth effect interacts with professions. *Psychological Research*. https://doi.org/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. https://doi.org/10.1038/srep18051

Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Schnall, S. (2011). Clean, Proper and Tidy Are More Than the Absence of Dirty, Disgusting and Wrong. *Emotion Review*, *3*(3), 264–266. https://doi.org/10.1177/1754073911402397

Schnall, S. (2014). Are there basic metaphors? In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life*. (pp. 225–247). American Psychological Association. http://content.apa.org/books/14278-010

Schnall, S. (2017). Disgust as embodied loss aversion. *European Review of Social Psychology*, *28*(1), 50–94. https://doi.org/10.1080/10463283.2016.1259844

*Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, *19*(12), 1219–1222. https://doi.org/10.1111/j.1467-9280.2008.02227.x

Schneider, I. K., Rutjens, B. T., Jostmann, N. B., & Lakens, D. (2011). Weighty Matters: Importance Literally Feels Heavy. *Social Psychological and Personality Science*, *2*(5), 474–478. https://doi.org/10.1177/1948550610397895

*Schwader, K. L. (2013). *The Potential of Physical Cleansing for Goal Disengagement: An Embodied Cognition Approach* [Ph.D., Yale University]. http://search.proquest.com/psycinfo/docview/1496071625/abstract/90D5213A84584877PQ/1

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Erlbaum.

*Sekulak, M., & Maciuszek, J. (2017). Metaphorical Association Between Physical and Moral Purity in the Context of One's Own Transgressions and Immoral Behavior of Others. *Psychology of Language and Communication, 21*(1), 152–170. https://doi.org/10.1515/plc-2017-0008

Senn, S. J. (2009). Overstating the evidence – double counting in meta-analysis and related problems. *BMC Medical Research Methodology*, *9*(1), 10. https://doi.org/10.1186/1471-2288-9-10

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, j4008. https://doi.org/10.1136/bmj.j4008

*Sheikh, S., Botindari, L., & White, E. (2013). Embodied Metaphors and Emotions in the Moralization of Restrained Eating Practices. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2012.12.016

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change* (pp. xii, 218). Yale Univer. Press.

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

Siev, J., Zuckerman, S. E., & Siev, J. J. (2018). The Relationship Between Immorality and Cleansing: A Meta-Analysis of the Macbeth Effect. *Social Psychology*, 1–7. https://doi.org/10.1027/1864-9335/a000349

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U. (2014, March 12). *[17] No-way Interactions*. Data Colada. http://datacolada.org/17

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). *p* -Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science*, *9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2014, December 3). *[30] Trim-and-Fill is Full of It (bias)*. Data Colada. https://datacolada.org/30

*Sobol, K., & Giroux, M. (2021). Threat specificity in fear appeals and the role of perceived assimilation to threat. *Advances in Consumer Research*, *49*, 439–440. https://www.acrwebsite.org/volumes/3000867/volumes/v49/NA-49

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845

Spruijt, B. M., van Hooff, J. A., & Gispen, W. H. (1992). Ethology and neurobiology of grooming behavior. *Physiological Reviews*, *72*(3), 825–852.

Stanford Medicine. (2022). *Symptoms*. Obsessive-Compulsive and Related Disorders. https://med.stanford.edu/ocd/about/symptoms.html

Stanley, T. D. (2017). Limitations of PET-PEESE and Other Meta-Analysis Methods. *Social Psychological and Personality Science*, *8*(5), 581–591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Stanley, T. D., & Doucouliagos, H. (2015). Neither fixed nor random: Weighted least squares meta-analysis. *Statistics in Medicine*, *34*(13), 2116–2127. https://doi.org/10.1002/sim.6481

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*(1), 19–42. https://doi.org/10.1002/jrsm.1211

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. A. (2017). Finding the power to reduce publication bias: Finding the power to reduce publication bias. *Statistics in Medicine*. https://doi.org/10.1002/sim.7228

Sterne, J. A. C., & Egger, M. (2006). Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis* (pp. 99–110). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470870168.ch6

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis. *Journal of Clinical Epidemiology*, *53*(11), 1119–1129. https://doi.org/10.1016/S0895-4356(00)00242-0

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*(jul22 1), d4002–d4002. https://doi.org/10.1136/bmj.d4002

*Strachan, E., Schimel, J., Arndt, J., Williams, T., Solomon, S., Pyszczynski, T., & Greenberg, J. (2007). Terror Mismanagement: Evidence That Mortality Salience Exacerbates Phobic and Compulsive Behaviors. *Personality and Social Psychology Bulletin*, *33*(8), 1137–1151. https://doi.org/10.1177/0146167207303018

Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology*, *23*(1), 53–62. https://doi.org/10.1002/ejsp.2420230105

Strohminger, N. (2014). Disgust Talked About. *Philosophy Compass*, *9*(7), 478–493. https://doi.org/10.1111/phc3.12137

*Tang, H., Lu, X., Su, R., Liang, Z., Mai, X., & Liu, C. (2017). Washing away your sins in the brain: Physical cleaning and priming of cleaning recruit different brain networks after moral threat. *Social Cognitive and Affective Neuroscience*, *12*(7), 1149–1158. https://doi.org/10.1093/scan/nsx036

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: A Tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

Taylor, S. E. (1998). The social being in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., 1–2, pp. 58–95). McGraw-Hill.

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

The World Bank. (2011). *India—National Ganga River Basin Project* (61625; pp. 1–175). The World Bank. http://documents.worldbank.org/curated/en/710791468269101705/India-National-Ganga-River-Basin-Project

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How Linguistic Metaphor Scaffolds Reasoning. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2017.07.001

*Thomas, J., Grey, I., & Al-Romaithi, S. (2018). Macbeth Arabia: Moral threat and cleansing-related construct accessibility in Arab women. *Mental Health, Religion & Culture*, *21*(2), 131–138. https://doi.org/10.1080/13674676.2017.1381675

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

*Tobia, K., Chapman, G., & Stich, S. (2013). Cleanliness is next to morality, even for philosophers. *Journal of Consciousness Studies*, *20*(11–12), 194–204.

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, *120*(1), 65–84. https://doi.org/10.1037/a0030778

U.S. Bureau of Labor Statistics. (2019). *American Time Use Survey*. https://www.bls.gov/tus/

U.S. Bureau of Labor Statistics. (2022). *American Time Use Survey*. https://www.bls.gov/tus/

Valliant, M. (2012, March 22). *Do Juice Cleanses Work? 10 Truths About The Fad*. HuffPost. https://www.huffpost.com/archive/ca/entry/do-juice-cleanses-work_n_1372305

van Aert, R. C. M. (2022). *puniform: Meta-Analysis Methods Correcting for Publication Bias* (0.2.5) [Computer software]. https://CRAN.R-project.org/package=puniform

van Aert, R. C. M., & van Assen, M. A. L. M. (2018). *Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method* [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/zqjr9

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting Meta-Analyses Based on *p* Values: Reservations and Recommendations for Applying *p* -Uniform and *p* -Curve. *Perspectives on Psychological Science*, *11*(5), 713–729. https://doi.org/10.1177/1745691616650874

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. https://doi.org/10.1037/met0000025

Vevea, J. L., & Woods, C. M. (2005). Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the **metafor** Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

*Waller, K., & Boschen, M. J. (2015). Evoking and reducing mental contamination in female perpetrators of an imagined non-consensual kiss. *Journal of Behavior Therapy and Experimental Psychiatry, 49*, 195–202. https://doi.org/10.1016/j.jbtep.2014.07.009

Wang, M. C., & Bushman, B. J. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*, *3*(1), 46–54. https://doi.org/10.1037/1082-989X.3.1.46

Wang, Y., Gong, Y., & Li, Y. (2019). 不道德传染的心理机制及其影响 [The mental mechanism of immoral contagion and its effects]. *心理科学进展 [Advances in Psychological Science]*, *27*(4), 1–11.

*Ward, S. J., & King, L. A. (2015). Individual differences in intuitive processing moderate responses to moral transgressions. *Personality and Individual Differences, 87*, 230–235. https://doi.org/10.1016/j.paid.2015.08.021

Weber, M. (1919). *Politik als Beruf (Politics as a profession)* (1st ed.). Duncker & Humblot.

Weiss, A., Forstmann, M., & Burgmer, P. (2021). Moralizing mental states: The role of trait self-control and control perceptions. *Cognition*, *214*, 104662. https://doi.org/10.1016/j.cognition.2021.104662

Wesley, J. (1778). *Cleanliness is indeed next to godliness*.
West, C., & Zhong, C.-B. (2015). Moral Cleansing. *Current Opinion in Psychology*, *6*, 221–225. https://doi.org/10.1016/j.copsyc.2015.09.022
Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, *139*(2), 271–299. https://doi.org/10.1037/a0031640
Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology*, *39*(7), 1257–1267. https://doi.org/10.1002/ejsp.665
*Winterich, K. P., Mittal, V., & Morales, A. C. (2014). Protect thyself: How affective self-protection increases self-interested, unethical behavior. *Organizational Behavior and Human Decision Processes, 125*(2), 151–161. https://doi.org/10.1016/j.obhdp.2014.07.004
Wyer, R. S., Xu, A. J., & Shen, H. (2012). The effects of past behavior on future goal-directed activity. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 46, pp. 237–283). Academic. http://linkinghub.elsevier.com/retrieve/pii/B9780123942814000143
*Xie, W., Yu, B., Zhou, X., Sedikides, C., & Vohs, K. D. (2014). Money, moral transgressions, and blame. *Journal of Consumer Psychology*, *24*(3), 299–306. https://doi.org/10.1016/j.jcps.2013.12.002
*Xu, A. J., Zwick, R., & Schwarz, N. (2012). Washing away your (good or bad) luck: Physical cleansing affects risk-taking behavior. *Journal of Experimental Psychology: General*, *141*(1), 26–30. https://doi.org/10.1037/a0023997
*Xu, H., Bègue, L., & Bushman, B. J. (2014). Washing the guilt away: Effects of personal versus vicarious cleansing on guilty feelings and prosocial behavior. *Frontiers in Human Neuroscience*, *8*, 97. https://doi.org/10.3389/fnhum.2014.00097
Yan, S.-C. (2011). 身体洁净与道德 [Physical Cleanliness and Morality]. *心理科学进展 [Advances in Psychological Science]*, *19*(8), 1242–1248.
*Yan, Z., Ding, D., & Yan, L. (2011). To Wash Your Body, or Purify Your Soul: Physical Cleansing Would Strengthen the Sense of High Moral Character. *Psychology, 02*(09), 992–997. https://doi.org/10.4236/psych.2011.29149
*Yang, Q., Wu, X., Zhou, X., Mead, N. L., Vohs, K. D., & Baumeister, R. F. (2013). Diverging Effects of Clean Versus Dirty Money on Attitudes, Values, and Interpersonal Behavior. *Journal of Personality and Social Psychology*, *104*(3), 473–489. https://doi.org/10.1037/a0030596
*Yelderman, L. A., Haggard, M., Olsen, K., Ahmadi, Y., Knighten, K., & Rowatt, W. (2012). *Disgust, morality, religion, and social attitudes* [Unpublished raw data].
Zerkel, E. (2013, July 1). *World's Most Polluted Rivers*. The Weather Channel. https://weather.com/news/news/worlds-most-polluted-rivers-20130627
*Zestcott, C. A., & Stone, J. (2017). *The Role of Conscious Attention in How Cleanliness Serves as an Embodiment of Morality* [Unpublished raw data]. The University of Arizona.
*Zhong, C.-B. (2007). *Unpublished raw data*.
*Zhong, C.-B., & House, J. (2012, January). *Once a criminal, always a criminal: How stain irreversibility affects recidivism judgments*. Annual meeting of the Society for Personality and Social Psychology, San Diego, CA.
*Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*(5792), 1451–1452. https://doi.org/10.1126/science.1130726
*Zhong, C.-B., Strejcek, B., & Sivanathan, N. (2010). A clean self can render harsh moral judgment. *Journal of Experimental Social Psychology*, *46*(5), 859–862. https://doi.org/10.1016/j.jesp.2010.04.003

Table 1

Expected Assimilation vs. Contrast Effect as a Function of the Polarity of Physical Cleansing and Its Associated Psychological Variable

If this pole of physical cleansing	is expected to be empirically linked to this pole of the psychological variable,	then the effect is expected to be a/an _____ effect
Clean	Positive (e.g., moral, religious)	Assimilation
	Negative (e.g., immoral, non-religious)	Contrast
Unclean	Positive (e.g., moral, religious)	Contrast
	Negative (e.g., immoral, non-religious)	Assimilation

Table 2

Synthesized Effect Size Estimates Using Three Statistical Models Without Excluding Outliers or After Excluding Outliers Based on Leave-One-Out Diagnoses or Sensitivity Analysis

Variable	No exclusion	Excluding OIC-RE	Sensitivity analysis: Manually excluding effects with g >				
			4.0	2.0	1.5	1.0	0.5
Count							
Study-level (k_{study})	230	230	229	229	229	217	162
Effect-level (k_{effect})	551	535	546	537	521	479	318
RE model							
Heterogeneity statistics							
τ^2	0.198	0.110	0.172	0.121	0.085	0.052	0.012
SE of τ^2	0.016	0.011	0.014	0.011	0.009	0.007	0.003
I^2	84.767	75.980	82.960	77.582	71.173	61.469	31.347
H^2	6.565	4.163	5.869	4.461	3.469	2.595	1.457
Homogeneity statistic (Q)							
Q	2669.709	1886.997	2408.293	1970.594	1600.540	1154.066	464.497
df of Q	550	534	545	536	520	478	317
p of Q	1.53E-274	3.64E-150	1.22E-231	9.47E-163	2.44E-110	8.768E-58	1.22E-07
Effect size estimate (g)							
g	0.455	0.393	0.437	0.402	0.362	0.302	0.143
SE of g	0.022	0.018	0.021	0.019	0.017	0.015	0.013
z	20.413	21.481	20.687	21.356	21.360	19.905	11.092
p of z	1.293E-92	2.32E-102	4.555E-95	3.46E-101	3.17E-101	3.679E-88	1.379E-28
LB of 95% CI of g	0.411	0.357	0.396	0.365	0.329	0.272	0.118
UB of 95% CI of g	0.498	0.429	0.479	0.439	0.396	0.332	0.169
RE model using RVE with SSC							
Effect size estimate (g)							
g	0.411	0.385	0.395	0.384	0.370	0.315	0.160
SE of g	0.024	0.022	0.023	0.022	0.021	0.020	0.018
t	16.777	17.440	16.989	17.318	17.500	16.049	8.975
df of t	210.160	204.960	206.920	204.554	202.281	185.854	115.973
p of t	1.219E-40	2.67E-42	4.777E-41	6.3E-42	2.417E-42	6.981E-37	6.33E-15
LB of 95% CI of g	0.362	0.342	0.349	0.340	0.329	0.277	0.124
UB of 95% CI of g	0.459	0.429	0.441	0.428	0.412	0.354	0.195
RE model using RVE without SSC							
Effect size estimate (g)							
g	0.411	0.385	0.395	0.384	0.370	0.315	0.160
SE of g	0.024	0.022	0.023	0.022	0.021	0.020	0.018
t	16.777	17.446	16.992	17.324	17.507	16.057	8.987
df of t	229	229	228	228	228	216	161
p of t	9.701E-42	6.309E-44	2.188E-42	1.813E-43	4.573E-44	1.041E-38	6.544E-16
LB of 95% CI of g	0.362	0.342	0.349	0.340	0.329	0.277	0.125
UB of 95% CI of g	0.459	0.429	0.441	0.428	0.412	0.354	0.195

Note. OIC-RE = outliers and influential cases diagnosed on the basis of the RE model. RE = random-effects. RVE = robust variance estimates. SSC = small-sample corrections. k_{study} = count of studies. k_{effect} = count of effects. τ^2 = amount of total heterogeneity between effects. *SE* = standard error. I^2 = percentage of total variability due to heterogeneity between effects. H^2 = ratio of total variability to sampling variability. Q = homogeneity statistic. g = Hedges' g. LB = lower bound. UB = upper bound. CI = confidence interval.

Table 3

Key Results of Assessing and Correcting for Publication Bias Using Eight Methods Without Excluding Outliers or After Excluding Outliers Based on Leave-One-Out Diagnoses or Sensitivity Analysis (Full Results in Table S5)

Variable	No exclusion	Excluding OIC-RE	Sensitivity analysis: Excluding effects with g > 4.0	2.0	1.5	1.0	0.5
Count of studies (k_{study})	230	230	229	229	229	217	162
Count of effects (k_{effect})	551	535	546	537	521	479	318
Method 1: Normal-quantile plot (Figures S6a-S6b and S6d-S6h) with Shapiro-Wilk normality test							
W	0.601***	0.844***	0.950***	0.985***	0.997	0.984***	0.929***
Method 2: Funnel plot (Figures S3a, S5a, S5e, S5g, S5i, S5k, and S5m) with Egger regression test							
Slope of Egger regression test	1.793***	1.476***	1.657***	1.499***	1.326***	1.059***	0.219
Method 3: WAAP-WLS[a]							
g in WLS	0.283***	0.263***	0.281***	0.267***	0.251***	0.220***	0.128***
g in WAAP (based on study-level *n*)	0.103***	0.107**	0.103***	0.107***	0.111**	0.103*	0.108
g in WAAP (based on subgroup-level *n*)	0.104**	0.107**	0.104**	0.107**	0.111**	0.103*	0.108
Method 4: PET-PEESE[b]							
g in PET	0.004	0.035	0.023	0.035	0.048*	0.063**	0.099***
g in PEESE	0.152***	0.163***	0.167***	0.166***	0.162***	0.152***	0.120***
g in PET-PEESE	0.004	0.035	0.023	0.035	0.162***	0.152***	0.120***
Method 5: *p*-uniform and *p*-uniform*							
g in *p*-uniform	0.522***	0.459***	0.503***	0.467***	0.414***	0.277***	-0.040
g in *p*-uniform*	0.472***	0.348***	0.440***	0.367***	0.298***	0.220***	0.116***
Method 6: Selection modeling							
g in 3PSM for RE model	0.451***	0.329***	0.423***	0.349***	0.282***	0.209***	0.125***
g in 4PSM for RE model	0.320***	0.260***	0.312***	0.274***	0.225***	0.164***	0.102***
Method 7: PET-PEESE[b] using RVE							
g in PET using RVE with SSC	-0.104	0.007	0.020	0.034	0.053	0.099*	0.156***
g in PEESE using RVE with SSC	0.178***	0.226***	0.240***	0.240***	0.240***	0.233***	0.169***
g in PET-PEESE using RVE with SSC	-0.104	0.007	0.020	0.034	0.053	0.233***	0.169***

Variable	No exclusion	Excluding OIC-RE	Sensitivity analysis: Excluding effects with g >				
			4.0	2.0	1.5	1.0	0.5
Method 8: Aggregating dependent effect sizes and submitting aggregated estimates to methods 1-6 (Table S7)							

Note. [a] When both WLS and WAAP estimates are available, the conditional estimator (WAAP-WLS) would recommend using the WAAP estimate. [b] When the PET estimate is statistically significant, the conditional estimator (PET-PEESE) would recommend using the PEESE estimate. g = Hedges' g. All p-values are two-tailed. * $p < .05$. ** $p < .01$. *** $p < .001$. OIC-RE = outliers and influential cases diagnosed on the basis of the RE model. RE = random-effects. WAAP = weighted average of adequately powered studies. WLS = weighted least squares meta-regression with no intercept. PET = precision-effect test, the statistical significance of which should be evaluated with a one-tailed alpha of .05 (Stanley & Doucouliagos, 2014). PEESE = precision-effect estimate with standard errors. PSM = parameter selection model, where the weight of each p-value interval was relative to the first p-value interval (one-tailed $p < .025$), which was fixed at 1.0 to address an indeterminancy (Vevea & Woods, 2005). RVE = robust variance estimates. SSC = small-sample corrections.

Table 4

Omnibus Significance Tests in Single-Moderator Analyses Based on the Random-Effects (RE) Meta-Regression, Precision-Effect Test (PET), and Precision-Effect Estimate with Standard Error (PEESE), All Using Robust Variance Estimates with Small-Sample Corrections

	Count		RE meta-regression				PET				PEESE			
Moderator	Study-level (k_{study})	Effect-level (k_{effect})	*F*	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*
Theoretical types														
Directionality	217	479	0.074	1	175.591	0.786	0.009	1	168.752	0.925	0.086	1	169.905	0.770
Theoretical basis	217	479	1.251	1	64.237	0.268	0.389	1	67.504	0.535	0.884	1	64.388	0.351
Content-based assimilation vs. contrast	171	401	0.496	1	149.409	0.482	0.759	1	137.425	0.385	0.990	1	139.977	0.321
Domains and subdomains														
Subdomain of physical cleansing	217	479	15.346	2	7.398	**0.002**	10.904	2	7.181	**0.007**	12.666	2	7.233	**0.004**
Domain of psychological variable: Relation to morality	217	479	1.761	2	77.368	0.179	1.692	2	76.154	0.191	2.051	2	76.580	0.136
If directly related to morality, which subdomain?	131	326	1.486	3	35.689	0.235	2.725	3	36.035	0.058	2.429	3	35.551	0.081
If in sanctity/degradation subdomain, sexual content?	37	92	0.263	2	10.205	0.774	0.117	2	9.826	0.891	0.106	2	10.029	0.900
If indirectly related or unrelated to morality, which subdomain?	101	184	0.135	1	84.174	0.714	0.669	1	74.600	0.416	0.456	1	77.495	0.502
Operationalizations of manipulation and measure														
Operationalization of manipulation	217	479	0.735	2	72.489	0.483	0.082	2	68.541	0.921	0.294	2	71.231	0.746
Operationalization of measure	217	479	5.495	2	69.252	**0.006**	2.646	2	61.973	0.079	3.240	2	65.197	**0.046**
Features of experimental and statistical design														
Mediation tested?	217	479	0.329	1	15.252	0.575	1.435	1	15.390	0.249	1.009	1	15.423	0.331
Mediation supported?	217	479	1.229	1	11.526	0.290	3.003	1	11.504	0.110	2.313	1	11.572	0.155
Was the measure a DV or a mediator?	217	479	0.284	1	8.273	0.608	1.052	1	8.254	0.334	0.836	1	8.288	0.386
Content-based specificity tested?	169	398	0.946	1	62.685	0.334	2.361	1	62.383	0.129	1.491	1	62.166	0.227
Content-based specificity supported?	169	398	4.814	1	57.129	**0.032**	5.663	1	51.984	**0.021**	5.678	1	54.501	**0.021**
Moderation tested?	217	479	3.198	1	152.649	0.076	6.759	1	130.460	**0.010**	5.013	1	139.691	**0.027**
Moderation supported?	217	479	0.033	1	100.472	0.857	0.154	1	85.396	0.696	0.038	1	90.034	0.846

	Count		RE meta-regression				PET				PEESE			
Moderator	**Study-level (k_{study})**	**Effect-level (k_{effect})**	***F***	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***
Did the focal factor include another condition?	217	479	0.468	1	37.088	0.498	0.834	1	36.081	0.367	0.648	1	36.230	0.426
Preregistration of study	217	479	12.537	1	10.797	**0.005**	5.878	1	12.870	**0.031**	8.459	1	11.742	**0.013**
Participant demographics														
Type of participants	217	479	0.976	2	38.111	0.386	4.897	2	37.830	**0.013**	3.056	2	37.502	0.059
Region of participants	217	479	3.822	3	24.020	**0.023**	1.961	3	23.836	0.147	2.478	3	23.422	0.086
Female percentage of participants	173	392	1.031	1	31.080	0.318	2.756	1	29.348	0.108	2.992	1	29.593	0.094
Report characteristics														
Type of report	217	479	9.721	1	39.558	**0.003**	10.162	1	37.222	**0.003**	9.812	1	38.113	**0.003**
How the report was presented by authors	217	479	60.844	2	13.189	**0.000**	45.003	2	13.031	**0.000**	53.309	2	12.969	**0.000**
Year of report	217	479	6.597	1	35.263	**0.015**	2.885	1	31.705	0.099	3.598	1	33.812	0.066

Note. *df1* = numerator degrees of freedom. *df2* = denominator degrees of freedom.

Table 5

Omnibus Significance Tests for Interactions between Directionality (i.e., Psychological Consequence vs. Antecedent of Physical Cleansing) and Other Moderators of Cleansing Effects Based on the Random-Effects (RE) Meta-Regression, Precision-Effect Test (PET), and Precision-Effect Estimate with Standard Error (PEESE), All Using Robust Variance Estimates with Small-Sample Corrections

	Count		RE meta-regression				PET				PEESE			
Moderator	**Study-level (k_{study})**	**Effect-level (k_{effect})**	***F***	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***
Theoretical types														
Theoretical basis	217	479	10.510	1	33.739	**0.003**	15.350	1	32.407	**0.000**	13.768	1	33.019	**0.001**
Content-based assimilation vs. contrast	171	401	4.073	1	73.007	**0.047**	5.152	1	69.194	**0.026**	5.183	1	71.550	**0.026**
Domains and subdomains														
Domain of psychological variable: Relation to morality	217	479	3.258	2	56.163	**0.046**	5.110	2	55.343	**0.009**	4.631	2	55.696	**0.014**
If directly related to morality, which subdomain?	131	326	4.851	3	34.335	**0.006**	3.859	3	33.505	**0.018**	4.031	3	33.845	**0.015**
If indirectly related or unrelated to morality, which subdomain?	101	184	0.029	1	41.880	0.864	0.005	1	41.983	0.945	0.001	1	43.011	0.982
Operationalizations of manipulation and measure														
Operationalization of manipulation	217	479	2.719	2	52.633	0.075	5.143	2	53.168	**0.009**	4.030	2	53.243	**0.023**
Operationalization of measure	217	479	2.937	2	65.781	0.060	2.246	2	64.454	0.114	2.776	2	64.467	0.070
Features of experimental and statistical design														
Mediation tested?	217	479	2.217	1	11.530	0.163	2.135	1	11.367	0.171	2.017	1	11.432	0.182
Mediation supported?	217	479	4.464	1	6.627	0.075	4.096	1	6.485	0.086	4.190	1	6.526	0.083
Was the measure a DV or a mediator?	217	479	1.382	1	1.867	0.368	0.732	1	1.899	0.487	0.804	1	1.890	0.469
Content-based specificity tested?	169	398	0.313	1	60.397	0.578	0.067	1	58.628	0.797	0.023	1	59.966	0.880
Content-based specificity supported?	169	398	0.573	1	54.816	0.452	0.001	1	54.415	0.980	0.102	1	54.582	0.751
Moderation tested?	217	479	0.741	1	120.369	0.391	1.841	1	116.543	0.178	1.341	1	118.921	0.249
Moderation supported?	217	479	0.074	1	94.830	0.787	0.056	1	89.772	0.814	0.017	1	92.397	0.898
Did the focal factor include another condition?	217	479	1.207	1	31.814	0.280	5.070	1	30.287	**0.032**	3.075	1	30.409	0.090
Preregistration of study	217	479	0.033	1	13.970	0.859	0.443	1	15.119	0.516	0.003	1	14.067	0.955

Moderator	Count		RE meta-regression				PET				PEESE			
	Study-level (k_{study})	Effect-level (k_{effect})	*F*	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*
Participant demographics														
Type of participants	217	479	5.877	2	16.893	**0.012**	2.545	2	16.697	0.108	3.684	2	16.594	**0.047**
Region of participants	217	479	1.584	3	13.796	0.238	1.168	3	13.407	0.358	1.425	3	13.419	0.279
Female percentage of participants	173	392	7.350	1	29.224	**0.011**	3.204	1	26.685	0.085	3.692	1	27.025	0.065
Report characteristics														
Type of report	217	479	0.530	1	11.580	0.481	1.059	1	10.623	0.326	1.183	1	10.674	0.301
How the report was presented by authors	217	479	0.189	2	2.404	0.839	0.507	2	2.489	0.653	0.268	2	2.504	0.784
Year of report	217	479	0.022	1	54.113	0.882	0.273	1	57.040	0.604	0.021	1	55.617	0.886

Note. *df1* = numerator degrees of freedom. *df2* = denominator degrees of freedom.

Table 6

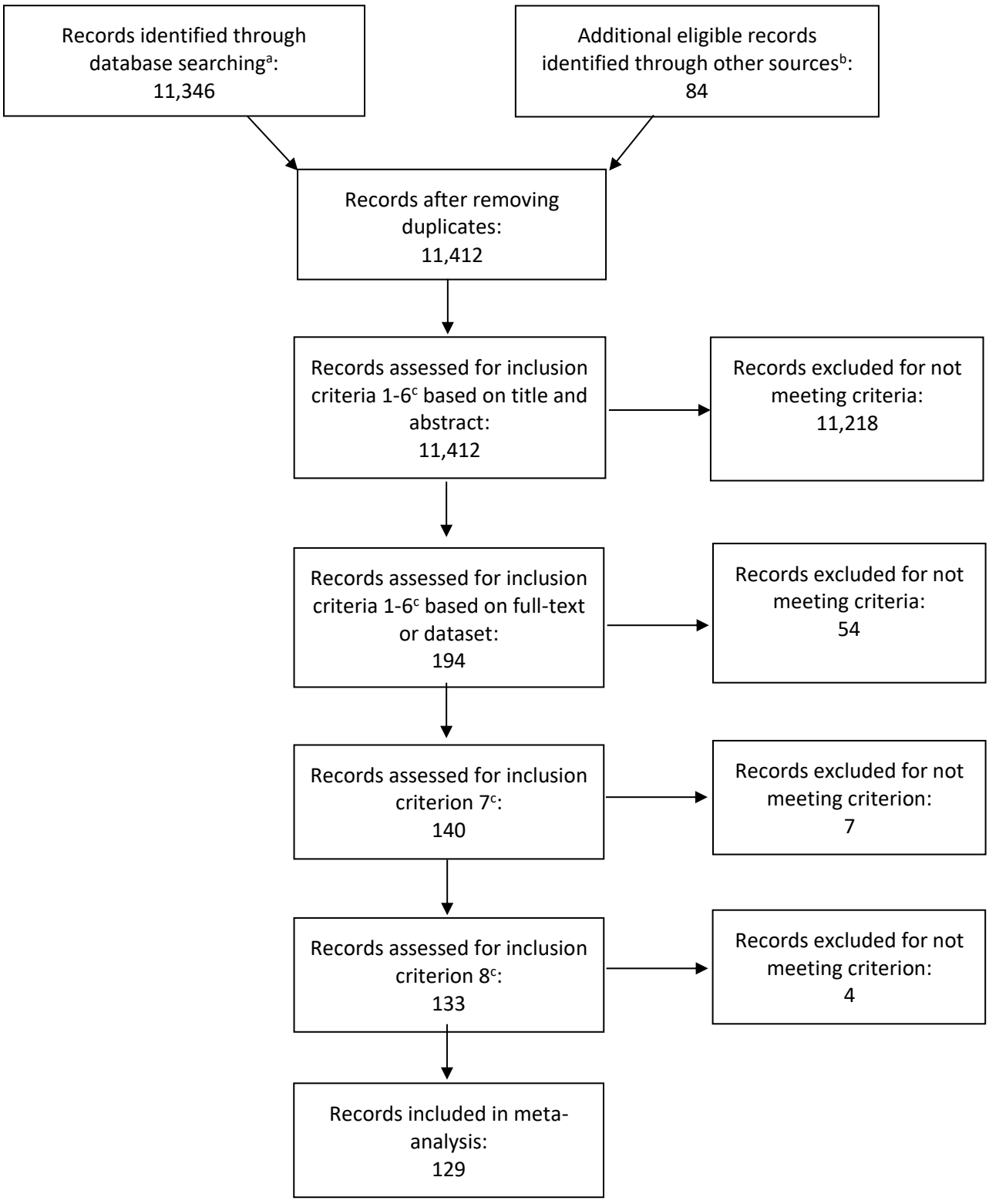
Consideration of Sample Size, Effect Size, and Statistical Power

			Descriptive statistics of study sample size						Synthesized effect size *d*		Power of "typical experiment" to detect effect size *d* with study sample size = *Mdn*, alpha = .05		Study sample size required for "typical experiment" to detect effect size *d* with power = .80, alpha = .05	
Set of experiments	k_{study}	k_{effect}	***Min***	**1st *Q***	***Mdn***	**3rd *Q***	***Max***	***M***	**RE model**	**FE model**	**RE model**	**FE model**	**RE model**	**FE model**
Including sample size outliers (*N* > 1,000)														
All experiments	200	479	16	58	86.5	148.5	7,001	200	0.306	0.221	0.290	0.174	338	643
Original experiments	166	401	16	57.25	84	136	4,142	168.2	0.361	0.301	0.372	0.275	243	349
All replications	34	78	28	61	120.5	209.5	7,001	355.3	0.060	0.037	0.062	0.055	8,704	22,698
Successful replications	8	29	28	55	68.5	134.8	222	101.5	0.217	0.217	0.143	0.143	670	670
Unsuccessful replications	26	49	28	72.25	135	209.5	7,001	433.4	-0.009	-0.009	0.050	0.050	430,321	430,321
Excluding sample size outliers (*N* > 1,000)														
All experiments	197	476	16	58	86	147	911	132.2	0.307	0.251	0.291	0.210	334	501
Original experiments	164	399	16	56.75	84	133	911	127.9	0.362	0.308	0.374	0.286	242	334
All replications	33	77	28	60	115	208	731	153.9	0.067	0.061	0.065	0.062	7,016	8,369
Successful replications	8	29	28	55	68.5	134.8	222	101.5	0.217	0.217	0.143	0.143	670	670
Unsuccessful replications	25	48	28	69	129	208	731	170.7	-0.017	-0.017	0.051	0.051	110,679	110,679

Note. *Min* = minimum, *Q* = quartile, *Mdn* = median, *Max* = maximum, *M* = mean. Effect size *d* is used here because it is the standard metric for power analysis in the *R* package *pwr* 1.3.0 (Champely et al., 2020). RE = random-effects. FE = fixed-effect. "Typical experiment" assumes a between-participant experimental design of one factor with two conditions. Actual experiments involved other designs (e.g., within-participant, more than one factor, more than two conditions in a factor), whose statistical power and sample sizes required for .80 power would be different from those in the table.

Figure 1

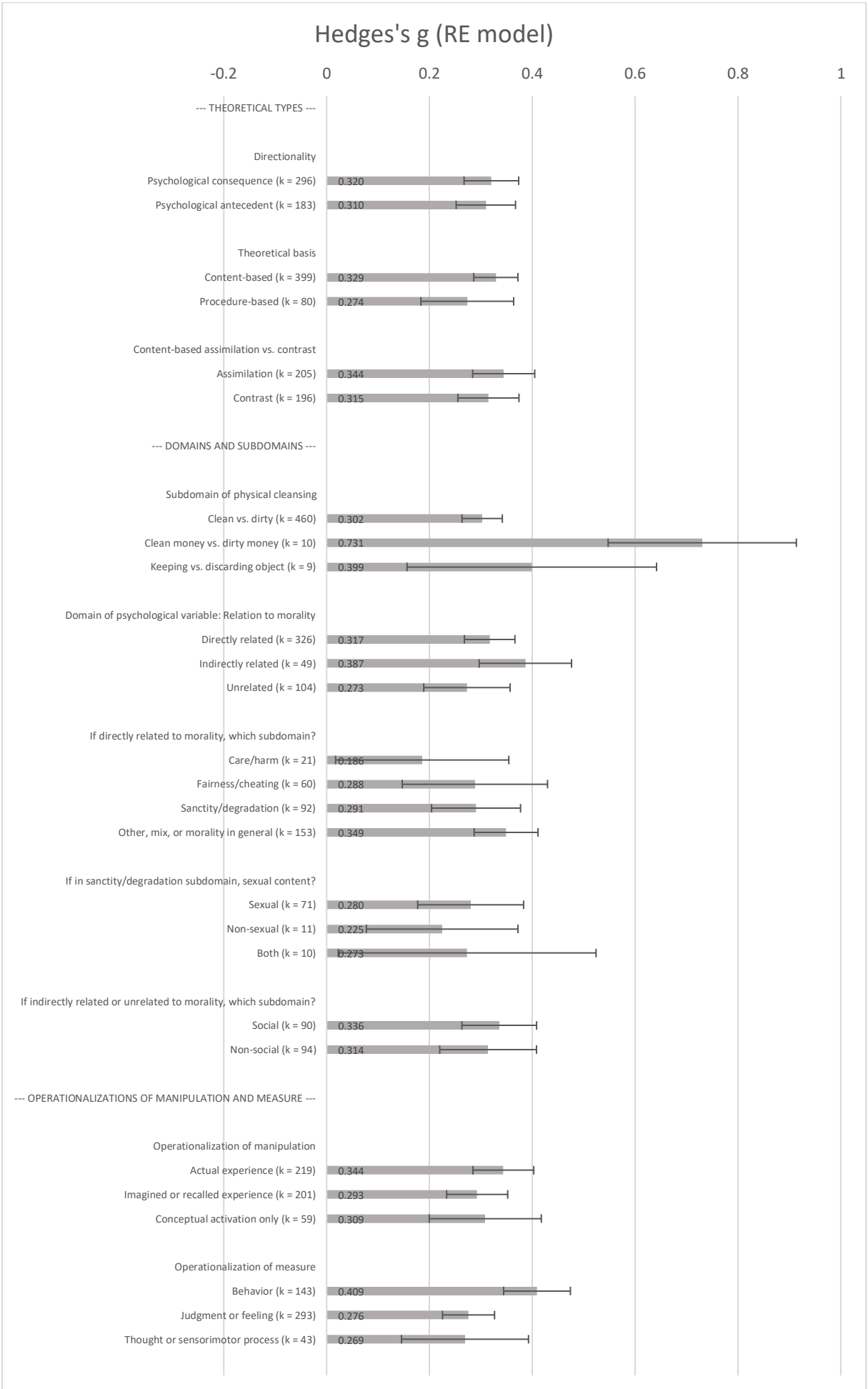
PRISMA-Style Flowchart Detailing Identification, Screening, Exclusion, and Inclusion of Records in the Present Meta-Analysis

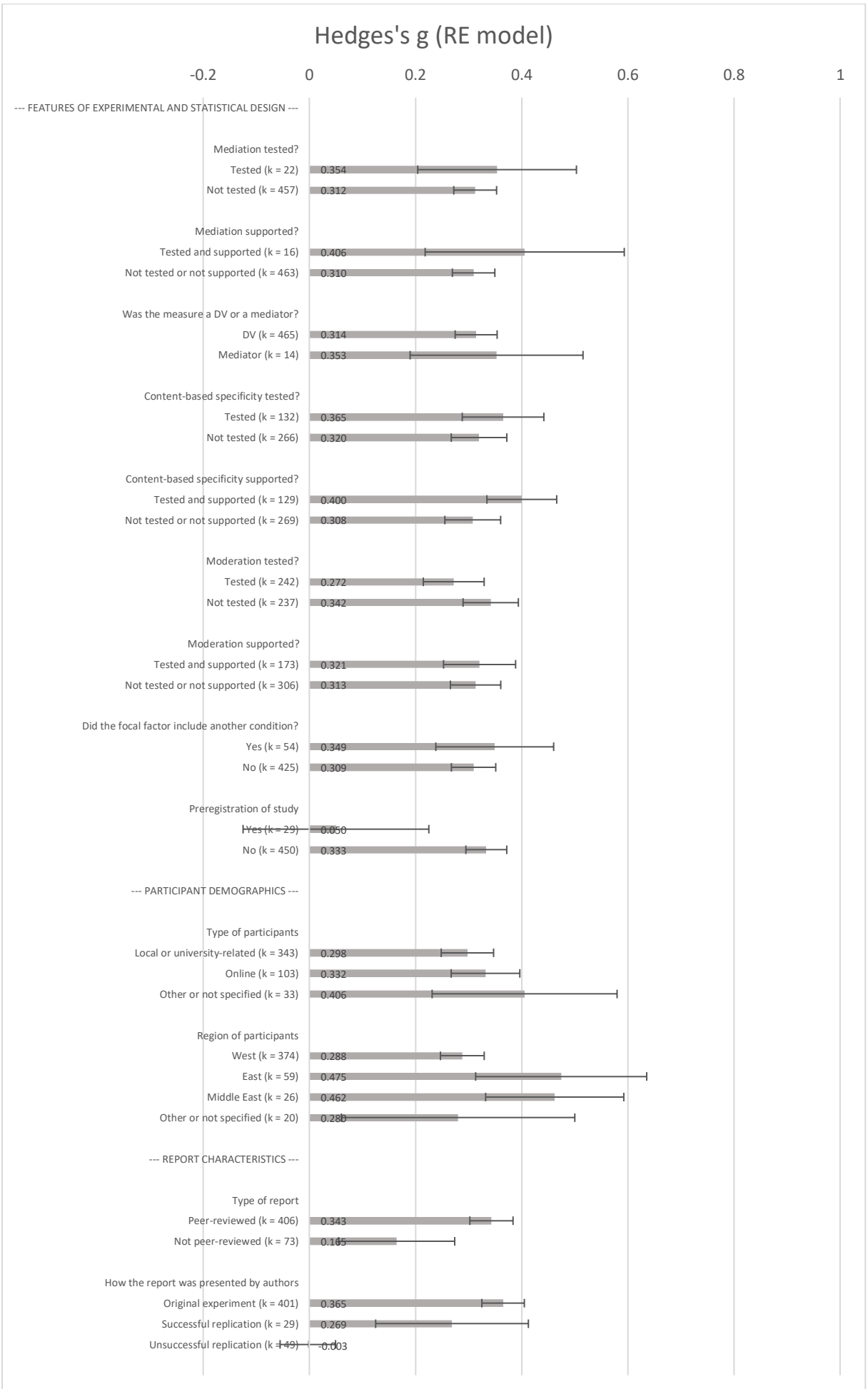


[a] We used the database APA PsycInfo. [b] Other sources included conference proceedings, PsychFileDrawer.org, recent large-scale replication projects, and listserv requests. [c] See main text (**Inclusion Criteria**) for full descriptions of inclusion criteria 1-8.

Figure 2

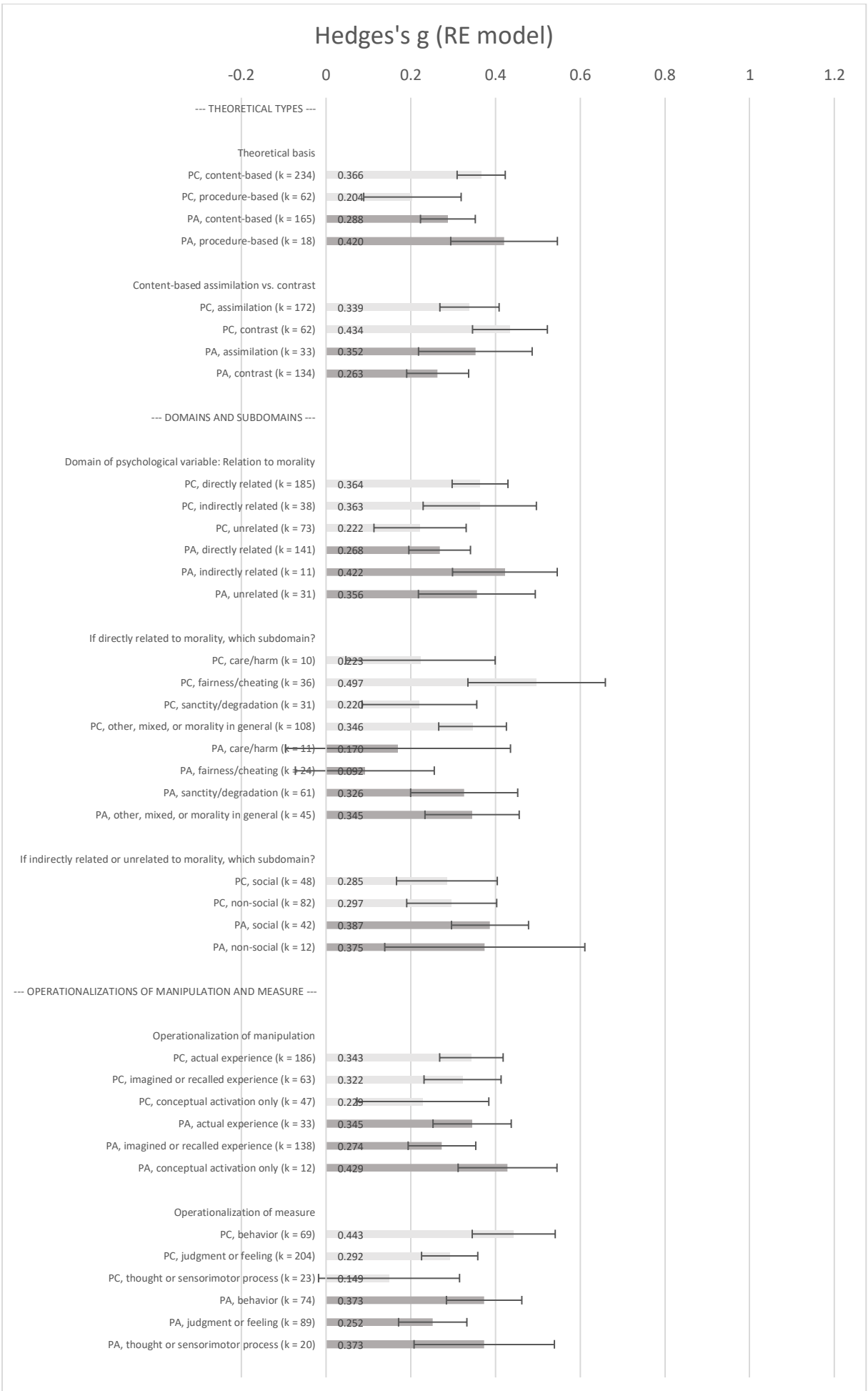
Single-Moderator Analyses Based on Random-Effects Meta-Regression Using Robust Variance Estimates with Small-Sample Corrections

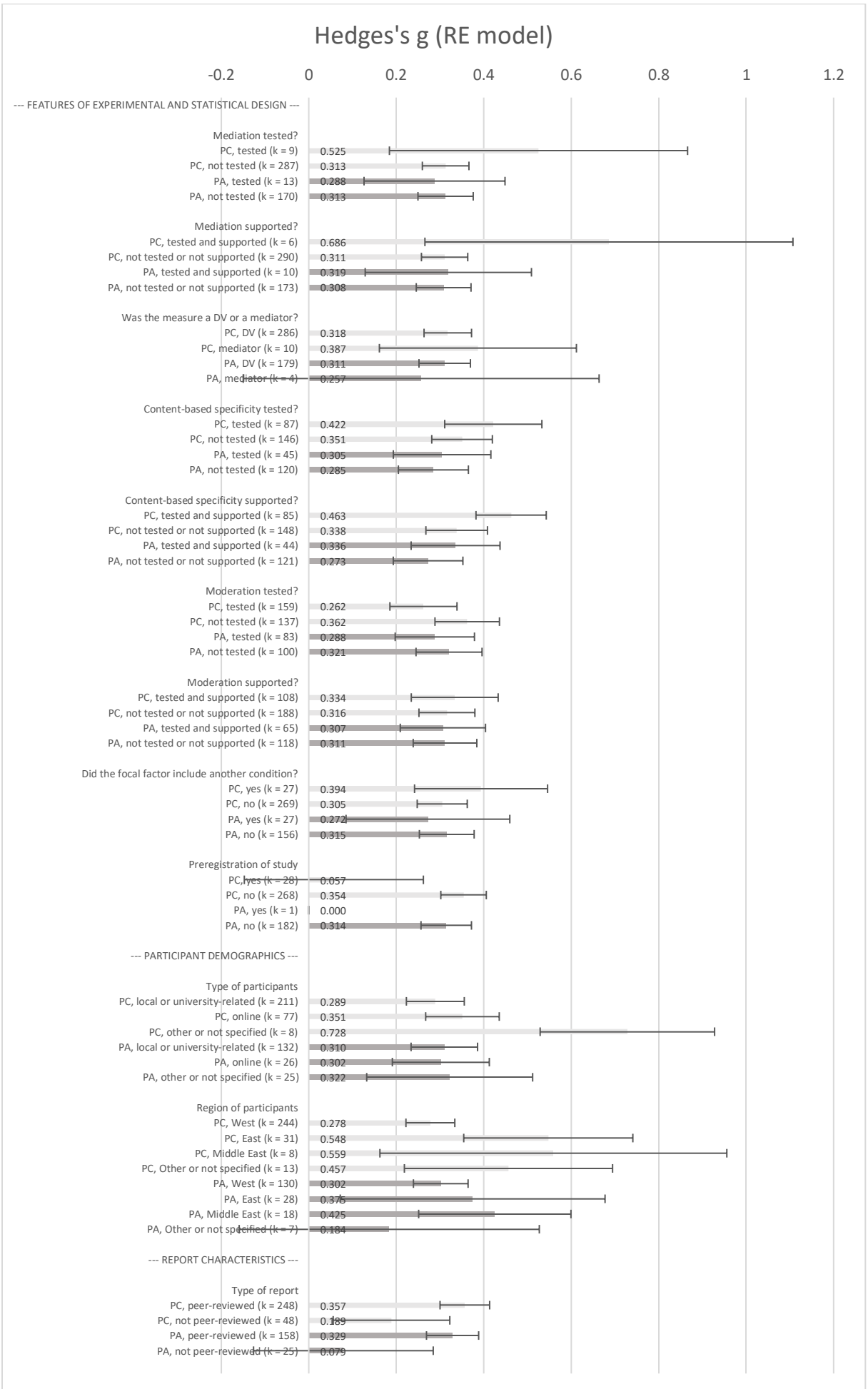

Hedges's g (RE model)
-0.2
0
0.2
0.4
0.6
0.8
1
--- THEORETICAL TYPES ---
Directionality
Psychological consequence (k = 296)
0.320
Psychological antecedent (k = 183)
0.310
Theoretical basis
Content-based (k = 399)
0.329
Procedure-based (k = 80)
0.274
Content-based assimilation vs. contrast
Assimilation (k = 205)
0.344
Contrast (k = 196)
0.315
--- DOMAINS AND SUBDOMAINS ---
Subdomain of physical cleansing
Clean vs. dirty (k = 460)
0.302
Clean money vs. dirty money (k = 10)
0.731
Keeping vs. discarding object (k = 9)
0.399
Domain of psychological variable: Relation to morality
Directly related (k = 326)
0.317
Indirectly related (k = 49)
0.387
Unrelated (k = 104)
0.273
If directly related to morality, which subdomain?
Care/harm (k = 21)
0.186
Fairness/cheating (k = 60)
0.288
Sanctity/degradation (k = 92)
0.291
Other, mix, or morality in general (k = 153)
0.349
If in sanctity/degradation subdomain, sexual content?
Sexual (k = 71)
0.280
Non-sexual (k = 11)
0.225
Both (k = 10)
0.273
If indirectly related or unrelated to morality, which subdomain?
Social (k = 90)
0.336
Non-social (k = 94)
0.314
--- OPERATIONALIZATIONS OF MANIPULATION AND MEASURE ---
Operationalization of manipulation
Actual experience (k = 219)
0.344
Imagined or recalled experience (k = 201)
0.293
Conceptual activation only (k = 59)
0.309
Operationalization of measure
Behavior (k = 143)
0.409
Judgment or feeling (k = 293)
0.276
Thought or sensorimotor process (k = 43)
0.269


Hedges's g (RE model)
-0.2
0
0.2
0.4
0.6
0.8
1
--- FEATURES OF EXPERIMENTAL AND STATISTICAL DESIGN ---
Mediation tested?
Tested (k = 22)
0.354
Not tested (k = 457)
0.312
Mediation supported?
Tested and supported (k = 16)
0.406
Not tested or not supported (k = 463)
0.310
Was the measure a DV or a mediator?
DV (k = 465)
0.314
Mediator (k = 14)
0.353
Content-based specificity tested?
Tested (k = 132)
0.365
Not tested (k = 266)
0.320
Content-based specificity supported?
Tested and supported (k = 129)
0.400
Not tested or not supported (k = 269)
0.308
Moderation tested?
Tested (k = 242)
0.272
Not tested (k = 237)
0.342
Moderation supported?
Tested and supported (k = 173)
0.321
Not tested or not supported (k = 306)
0.313
Did the focal factor include another condition?
Yes (k = 54)
0.349
No (k = 425)
0.309
Preregistration of study
Yes (k = 29)
0.050
No (k = 450)
0.333
--- PARTICIPANT DEMOGRAPHICS ---
Type of participants
Local or university-related (k = 343)
0.298
Online (k = 103)
0.332
Other or not specified (k = 33)
0.406
Region of participants
West (k = 374)
0.288
East (k = 59)
0.475
Middle East (k = 26)
0.462
Other or not specified (k = 20)
0.280
--- REPORT CHARACTERISTICS ---
Type of report
Peer-reviewed (k = 406)
0.343
Not peer-reviewed (k = 73)
0.165
How the report was presented by authors
Original experiment (k = 401)
0.365
Successful replication (k = 29)
0.269
Unsuccessful replication (k = 49)
-0.003

Note. Error bars represent 95% confidence intervals.

Figure 3
Double-Moderator Analyses Based on Random-Effects Meta-Regression Using Robust Variance Estimates with Small-Sample Corrections

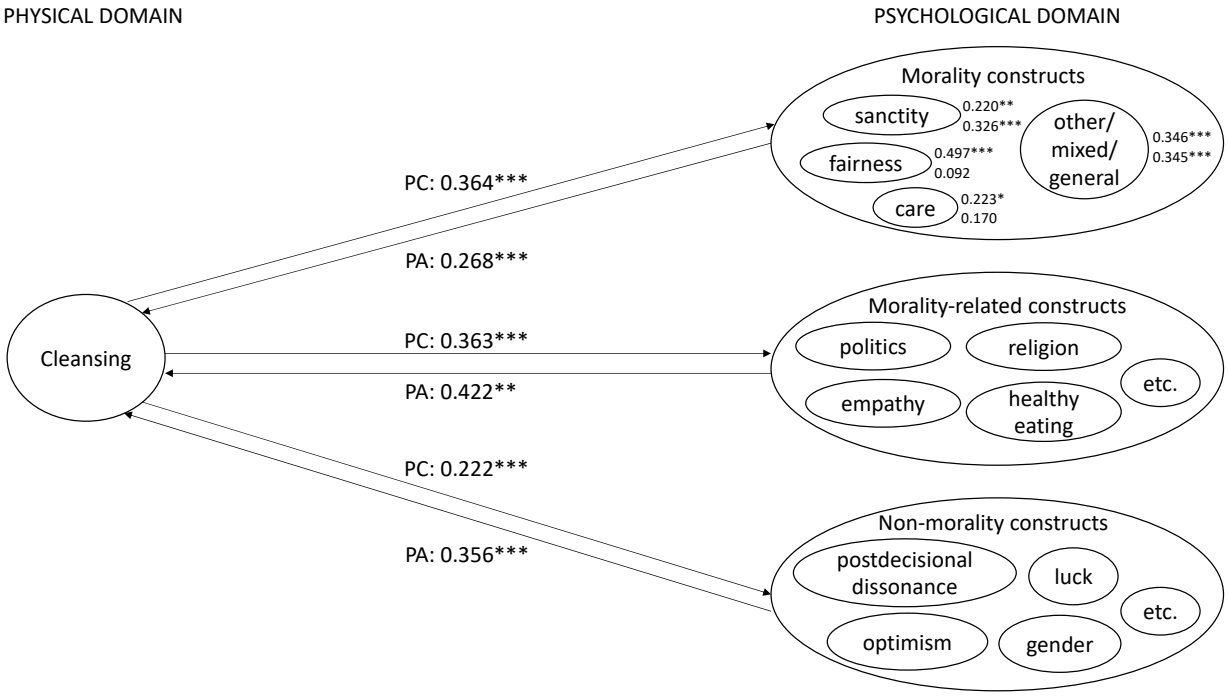

Hedges's g (RE model)
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
--- THEORETICAL TYPES ---
Theoretical basis
PC, content-based (k = 234)
0.366
PC, procedure-based (k = 62)
0.204
PA, content-based (k = 165)
0.288
PA, procedure-based (k = 18)
0.420
Content-based assimilation vs. contrast
PC, assimilation (k = 172)
0.339
PC, contrast (k = 62)
0.434
PA, assimilation (k = 33)
0.352
PA, contrast (k = 134)
0.263
--- DOMAINS AND SUBDOMAINS ---
Domain of psychological variable: Relation to morality
PC, directly related (k = 185)
0.364
PC, indirectly related (k = 38)
0.363
PC, unrelated (k = 73)
0.222
PA, directly related (k = 141)
0.268
PA, indirectly related (k = 11)
0.422
PA, unrelated (k = 31)
0.356
If directly related to morality, which subdomain?
PC, care/harm (k = 10)
0.223
PC, fairness/cheating (k = 36)
0.497
PC, sanctity/degradation (k = 31)
0.220
PC, other, mixed, or morality in general (k = 108)
0.346
PA, care/harm (k = 11)
0.170
PA, fairness/cheating (k = 24)
0.092
PA, sanctity/degradation (k = 61)
0.326
PA, other, mixed, or morality in general (k = 45)
0.345
If indirectly related or unrelated to morality, which subdomain?
PC, social (k = 48)
0.285
PC, non-social (k = 82)
0.297
PA, social (k = 42)
0.387
PA, non-social (k = 12)
0.375
--- OPERATIONALIZATIONS OF MANIPULATION AND MEASURE ---
Operationalization of manipulation
PC, actual experience (k = 186)
0.343
PC, imagined or recalled experience (k = 63)
0.322
PC, conceptual activation only (k = 47)
0.239
PA, actual experience (k = 33)
0.345
PA, imagined or recalled experience (k = 138)
0.274
PA, conceptual activation only (k = 12)
0.429
Operationalization of measure
PC, behavior (k = 69)
0.443
PC, judgment or feeling (k = 204)
0.292
PC, thought or sensorimotor process (k = 23)
0.149
PA, behavior (k = 74)
0.373
PA, judgment or feeling (k = 89)
0.252
PA, thought or sensorimotor process (k = 20)
0.373


Hedges's g (RE model)
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
--- FEATURES OF EXPERIMENTAL AND STATISTICAL DESIGN ---
Mediation tested?
PC, tested (k = 9) 0.525
PC, not tested (k = 287) 0.313
PA, tested (k = 13) 0.288
PA, not tested (k = 170) 0.313
Mediation supported?
PC, tested and supported (k = 6) 0.686
PC, not tested or not supported (k = 290) 0.311
PA, tested and supported (k = 10) 0.319
PA, not tested or not supported (k = 173) 0.308
Was the measure a DV or a mediator?
PC, DV (k = 286) 0.318
PC, mediator (k = 10) 0.387
PA, DV (k = 179) 0.311
PA, mediator (k = 4) 0.257
Content-based specificity tested?
PC, tested (k = 87) 0.422
PC, not tested (k = 146) 0.351
PA, tested (k = 45) 0.305
PA, not tested (k = 120) 0.285
Content-based specificity supported?
PC, tested and supported (k = 85) 0.463
PC, not tested or not supported (k = 148) 0.338
PA, tested and supported (k = 44) 0.336
PA, not tested or not supported (k = 121) 0.273
Moderation tested?
PC, tested (k = 159) 0.262
PC, not tested (k = 137) 0.362
PA, tested (k = 83) 0.288
PA, not tested (k = 100) 0.321
Moderation supported?
PC, tested and supported (k = 108) 0.334
PC, not tested or not supported (k = 188) 0.316
PA, tested and supported (k = 65) 0.307
PA, not tested or not supported (k = 118) 0.311
Did the focal factor include another condition?
PC, yes (k = 27) 0.394
PC, no (k = 269) 0.305
PA, yes (k = 27) 0.272
PA, no (k = 156) 0.315
Preregistration of study
PC, yes (k = 28) 0.057
PC, no (k = 268) 0.354
PA, yes (k = 1) 0.000
PA, no (k = 182) 0.314
--- PARTICIPANT DEMOGRAPHICS ---
Type of participants
PC, local or university-related (k = 211) 0.289
PC, online (k = 77) 0.351
PC, other or not specified (k = 8) 0.728
PA, local or university-related (k = 132) 0.310
PA, online (k = 26) 0.302
PA, other or not specified (k = 25) 0.322
Region of participants
PC, West (k = 244) 0.278
PC, East (k = 31) 0.548
PC, Middle East (k = 8) 0.559
PC, Other or not specified (k = 13) 0.457
PA, West (k = 130) 0.302
PA, East (k = 28) 0.375
PA, Middle East (k = 18) 0.425
PA, Other or not specified (k = 7) 0.184
--- REPORT CHARACTERISTICS ---
Type of report
PC, peer-reviewed (k = 248) 0.357
PC, not peer-reviewed (k = 48) 0.189
PA, peer-reviewed (k = 158) 0.329
PA, not peer-reviewed (k = 25) 0.079

Note. PC = psychological consequences of cleansing. PA = psychological antecedents of cleansing. Error bars represent 95% confidence intervals.

Figure 4

Bidirectionality and Domain-Generality of Cleansing Effects Based on Synthesized Effect Sizes (gs) in the Present Meta-Analysis



Note. PC = psychological consequences of cleansing. PA = psychological antecedents of cleansing. * $p < .05$. ** $p < .01$. *** $p < .001$. Morality constructs refer to psychological variables generally conceptualized as constituting or being directly about morality. Morality-related constructs refer to psychological variables generally conceptualized as partially overlapping with and thus being indirectly related to morality. Non-morality constructs refer to psychological variables generally conceptualized as not overlapping with and thus being unrelated to morality.