AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

2024, Vol. 150, No. 6, 694–726
https://doi.org/10.1037/bul0000441

Influences of Past Moral Behavior on Future Behavior: A Review of Sequential Moral Behavior Studies Using Meta-Analytic Techniques

Rose Ferguson[1, 2], Leah Kaufmann[2], Aimee Brown[3], and Xochitl de la Piedad Garcia[2]
[1] Institute of Health of Wellbeing, Federation University Australia
[2] School of Behavioural and Health Sciences, Australian Catholic University
[3] Turner Institute for Brain and Mental Health, Monash University

Experimental research on sequential moral behavior (SMB) has found that engaging in an initial moral (or immoral) behavior can sometimes lead to moral balancing (i.e., switching between positive and negative behavior) and sometimes to moral consistency (i.e., maintaining a consistent pattern of positive or negative behavior). In two meta-analyses, we present the first comprehensive syntheses of SMB studies and test moderators to identify the conditions under which moral balancing and moral consistency are most likely to occur. Meta-Analysis 1 ($k = 217$ effect sizes, $N = 31{,}242$) revealed that engaging in an initial positive behavior only reliably resulted in moral licensing (i.e., balancing) in studies that measured engagement in negative target behaviors (Hedges' $g = 0.25$, 95% CI [0.16, 0.44]) and only resulted in positive consistency in foot-in-the-door studies using prosocial requests (Hedges' $g = -0.44$, 95% CI [−0.59, −0.29]). Meta-Analysis 2 ($k = 132$ effect sizes, $N = 14{,}443$) revealed that engaging in an initial negative behavior only reliably resulted in moral compensation (i.e., balancing) in studies that measured engagement in positive target behaviors (Hedges' $g = 0.27$, 95% CI [0.18, 0.37]). We found no evidence for reliable negative consistency effects in any conditions. These results cannot be readily explained by current theories of SMB effects, and so further research is needed to better understand the mechanisms that drive moral balancing and consistency under the conditions observed.

Public Significance Statement
Previous research suggests the likelihood of engaging in a moral (or immoral) behavior may be influenced by whether a person has previously engaged in another moral action. Our analyses have identified different circumstances under which prior moral (or immoral) behavior leads to a greater likelihood of subsequent positive, and negative, behavior. These results may inform the design of future research that aims to better understand the mechanisms underlying sequential moral behavior effects. Understanding the dynamics of sequential moral behavior may also inform practical strategies to increase engagement in positive behavior, such as engagement with charitable causes.

Keywords: sequential behavior, moral behavior, moral licensing, moral consistency, moral compensation

A key goal for research in moral psychology is to identify factors that predict how a person is likely to act when given the opportunity to engage in a moral (or immoral) behavior. One factor that has become a particular area of interest, especially in the last few decades, is sequential moral behavior (SMB). Research investigating SMB aims to identify if engaging in an initial moral behavior influences participants' subsequent moral behavior. For this meta-analytic review, the term *moral behavior* will be used to refer broadly to behavior of any moral value (either positive or negative). The terms *positive* and *negative behavior* will be used to differentiate between moral behavior of different valence.

Rose Ferguson https://orcid.org/0000-0001-7630-4324
Leah Kaufmann https://orcid.org/0000-0003-0034-4068
Aimee Brown https://orcid.org/0000-0003-3072-8003
Xochitl de la Piedad Garcia https://orcid.org/0000-0002-9319-8671

The authors thank Kathryn Duncan for assisting with the development of our search strategy and Joel Anderson, Justin Timora, and Alex Poll for helpful feedback on the article.

Additional online material, data files, and analysis code can be found at the Open Science Framework repository (https://osf.io/upby2; Ferguson et al., 2023).

Open Access funding provided by Federation University Australia:

Rose Ferguson played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, and writing–original draft. Leah Kaufmann played a supporting role in methodology, supervision, and writing–review and editing. Aimee Brown played a supporting role in data curation. Xochitl de la Piedad Garcia played a lead role in supervision and a supporting role in conceptualization, methodology, project administration, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Rose Ferguson, Institute of Health of Wellbeing, Federation University Australia, Berwick, VIC 3806, Australia. Email: r.ferguson@federation.edu.au

Research on SMB uses an experimental design that has been termed the SMB paradigm (Mullen & Monin, 2016). The SMB paradigm involves presenting participants with a *moral manipulation* in which they are induced to perform, recall, or imagine engaging in a moral behavior. SMB studies utilize either positive (i.e., one based on positive behavior, e.g., recall helping another person) or negative manipulations (i.e., one based on negative behavior, e.g., recall harming another person). Participants are then presented with an opportunity to engage in a second moral behavior (referred to as the *target behavior*). If there are differences in engagement in the target behavior between participants in the moral manipulation and the control conditions, then it is assumed that the initial moral behavior has influenced engagement in the subsequent behavior. The SMB paradigm and different pattern of effects are depicted in Figure 1.

The results of research using the SMB paradigm have found evidence of two broad, conflicting patterns of behavior. The first, referred to as *moral balancing*, is characterized by oscillation between positive and negative behaviors. This can take the form of *moral licensing* (i.e., when an initial positive action leads to an increase in subsequent negative behavior; e.g., Monin & Miller, 2001) or *moral compensation* (i.e., when an initial negative manipulation leads to an increase in subsequent positive behavior; e.g., Jordan et al., 2011). The second, referred to as *moral consistency*, is characterized by a pattern of consistent moral behavior. This can take the form of *positive consistency* (i.e., when a positive manipulation leads to further engagement in positive behavior; e.g., Aquino et al., 2009) or *negative consistency* (i.e., when a negative manipulation leads to further negative behavior; e.g., Lee et al., 2016).

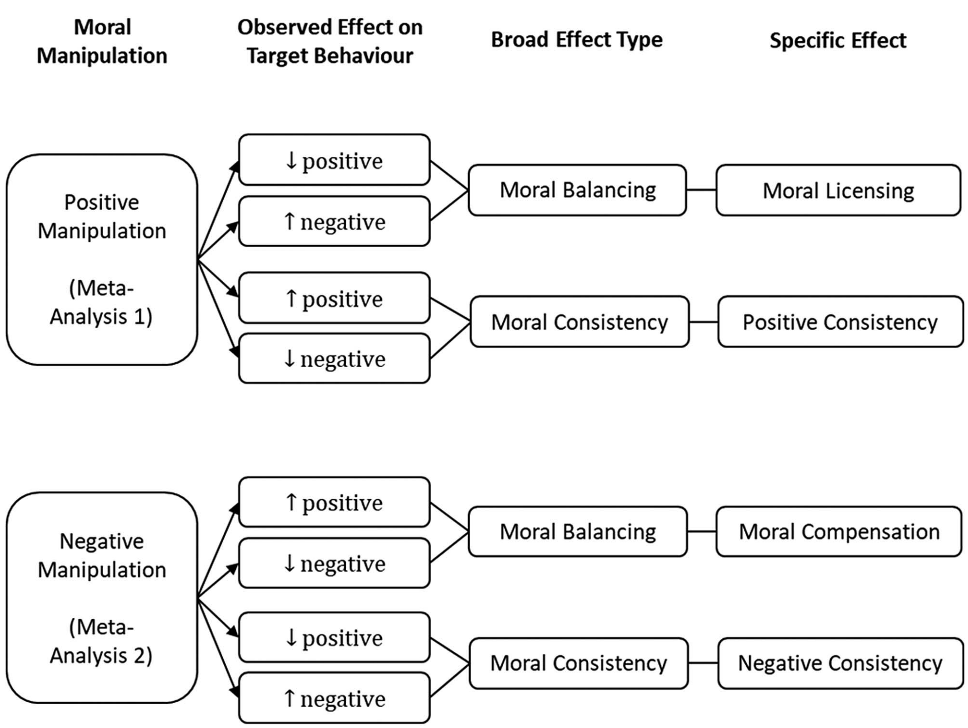
To date, research has not provided a clear answer to the question of when moral manipulations are likely to lead to moral balancing, and when they are likely to lead to moral consistency. A range of theories have been proposed that outline the conditions under which specific effects are likely to occur (e.g., moral credits, moral credentials, self-perception theory, mitigation of guilt; Burger, 1999; de Hooge, 2013; Lee et al., 2016; Monin & Miller, 2001; Nisan, 1990; Zhong et al., 2010). Each of these theories aims to explain only specific types of SMB effects (e.g., just moral balancing, specific types of moral balancing, or moral consistency) and so is not sufficient on their own to explain the full range of results. Researchers have argued that different SMB effects may be driven by different mechanisms, and so integrating theories may be the best approach to fully understanding the dynamic nature of SMB (e.g., Mullen & Monin, 2016). Before the proposed mechanisms for how SMB effects occur can be integrated, we must first identify the conditions under which each effect type occurs, and whether these conditions align with the predictions of these existing theories.

The Current Research

In this article, we present two meta-analyses of studies using the SMB paradigm. In Meta-Analysis 1, we present a synthesis of SMB studies using positive moral manipulations (i.e., moral licensing vs. positive consistency). In Meta-Analysis 2, we present a synthesis of SMB studies using negative moral manipulations (i.e., moral compensation vs. negative consistency). In these meta-analyses, we aim to identify methodological conditions under which

Figure 1

Variations of the Sequential Moral Behavior Paradigm and Possible Effects



Note. Upward (downward) facing arrows indicate an increase (decrease) in the target behavior relative to a control group.

moral balancing and moral consistency are most likely to occur and to examine whether these conditions align with predictions of existing theories of SMB effects. For each meta-analysis, we provide a brief review of relevant literature and an overview of current theoretical explanations for observed SMB effects. We then outline methodological moderators derived from these theories that can be tested using meta-analytic methods. Next, we present the method and results of the analyses and a discussion of these results. We end with a general discussion and recommendations for future research.

Meta-Analysis 1: Positive Moral Manipulations (Moral Licensing vs. Positive Consistency)

As previously discussed, SMB studies using positive manipulations (top half of Figure 1) can lead to moral licensing, positive consistency, or null results. Each pattern of results has typically been reported in different bodies of literature (e.g., studies self-described as moral licensing, moral consistency, foot-in-the-door [FITD]) and so we have arranged our review in line with this.

Literature Review

Moral Licensing Literature

Moral licensing describes a pattern of results such that engagement in a positive moral manipulation leads to greater engagement in subsequent morally negative behavior, compared to control (i.e., a switch from "good" behavior to "bad"). Moral licensing can take the form of either greater engagement in a negative target behavior (e.g., more cheating; Mazar & Zhong, 2010) or lower engagement in a positive target behavior (e.g., lower charitable donations; Sachdeva et al., 2009). Moral licensing came to prominence with seminal research by Monin and Miller (2001). Across three studies, they found that participants who were given the opportunity to demonstrate moral behavior by rejecting overtly prejudiced statements were subsequently more likely to express discriminatory attitudes in a subsequent hypothetical job hiring task, compared to comparison groups. The result of their research has since been replicated in a large-scale multisite Many Labs replication project (Ebersole et al., 2016), although the magnitude of the replicated effect (Cohen's $d = 0.15$) was notably smaller than that of the original study (Cohen's $d = 0.87$).

There has been a plethora of studies demonstrating moral licensing effects since the publication of Monin and Miller's (2001) seminal research. These effects have been found in a range of domains, such as pro-environmental behavior (e.g., Geng et al., 2016), volunteering (e.g., Kristofferson et al., 2013), cheating (Mazar & Zhong, 2010), economic games (e.g., Cornelissen et al., 2013), and self-indulgent choice (e.g., Khan & Dhar, 2006). Moral licensing has also been demonstrated in field studies (Karmarker & Bollinger, 2015; Tiefenbeck et al., 2013) and experience sampling research (Hofmann et al., 2014).

Despite the apparent wealth of research demonstrating moral licensing, there have also been several prominent failures to replicate moral licensing effects. Blanken et al. (2014) failed to replicate the results of Sachdeva et al. (2009), who originally found that participants who wrote stories about their moral traits (e.g., fair, generous) subsequently made lower charitable donations and were less likely to engage in hypothetical moral behavior than participants in a control group. Urban et al. (2019) failed to replicate the findings of Mazar and Zhong (2010), who found that participants who engaged in pro-environmental behavior subsequently cheated more than control participants. Finally, Rotella and Barclay (2020) failed to replicate the results of Conway and Peetz (2012), who found that participants who recalled a recent positive behavior subsequently made smaller charitable donations than participants who recalled a recent negative behavior. Indeed, Rotella and Barclay (2020) found participants who recalled a positive behavior donated *more* than both participants who recalled a negative behavior and participants who recalled a neutral behavior (a positive consistency effect). In addition to these failures to replicate existing effects, there are also several original studies that purport to test moral licensing that found no difference between positive moral manipulation and control conditions (e.g., Ghesla et al., 2019; Hayley & Zinkiewicz, 2013; Ho et al., 2015). These results call into question the ubiquity of moral licensing effects.

There have been two attempts to meta-analyze moral licensing effects (Blanken et al., 2015; Simbrunner & Schlegelmilch, 2017). Both analyses found a significant moral licensing effect when aggregating across all included studies (Cohen's $d = 0.31$, 95% CI [0.28, 0.38], $k = 91$ effect sizes, and Cohen's $d = 0.32$, 95% CI [0.23, 0.41], $k = 106$ effect sizes, respectively) that were not moderated by differences in methodology between studies. Simbrunner and Schlegelmilch (2017) found that effect sizes were moderated by culture, such that the moral licensing effect only occurred in North American and Western European samples, while positive consistency was observed for Southeast Asian samples (although this was based on a sample of only five studies). Neither meta-analysis identified methodological conditions that can explain when positive SMB studies are likely to lead to moral licensing, positive consistency, or null results.

The results of these past meta-analyses appear at face value to suggest that SMB studies using positive moral manipulations lead reliably to moral licensing, at least in Western samples. However, there are important limitations to these analyses that undermine such conclusions. Both analyses only included SMB studies, or conditions within SMB studies, in which the original study authors explicitly predicted moral licensing effects and excluded SMB studies using positive manipulations if the authors did not predict moral licensing. The decision to exclude these studies was not based on objective methodological criteria but on idiosyncratic predictions of the original study authors. In the absence of any universal, overarching theoretical explanation for when positive manipulations should lead to moral licensing, positive consistency, or null results, there is no theoretical justification for the exclusion of these studies. Furthermore, the bulk of evidence for positive consistency comes from studies that utilize SMB designs but do not label themselves as moral licensing studies and so were excluded from these analyses (e.g., FITD, spillover; e.g., Freedman & Fraser, 1966; Truelove et al., 2014). This means studies that are also valid tests of moral licensing but were less likely to find evidence of moral licensing were systematically excluded from these analyses, and thus the results represent only a subset of the SMB literature—a subset that is likely biased in favor of finding moral licensing.

The conclusions of both Blanken et al.'s (2015) and Simbrunner and Schlegelmilch's (2017) meta-analyses have also been challenged by Kuper and Bott (2019), who reanalyzed their data using more advanced approaches for detecting publication bias. Their results suggest that the average effect size for the included studies was likely much smaller than that estimated in the original meta-analyses and potentially no different to zero. Taking the results of Kuper and Bott's (2019) analyses, the failures to replicate prominent moral licensing effects, and original studies reporting null results, this leads to some doubt about the robustness of moral licensing and raises questions about where and when such effects are indeed likely to occur.

Positive Consistency Literature

Prosocial-FITD Studies. The most prominent source of evidence for positive consistency comes from the FITD literature. FITD is a compliance technique that involves presenting participants with an initial request that is small enough to elicit compliance (the experimental manipulation) and then measuring compliance with a subsequent larger request (e.g., Freedman & Fraser, 1966). Not all FITD studies meet all the definitional criteria of the SMB paradigm. Some FITD studies, which we refer to as *prosocial*-FITD studies, use requests for prosocial behavior such as helping, volunteering, or donating to charity (Arnold & Kaiser, 2016; Burger, 1999; Chartrand et al., 1999; Fointiat, 2006). These studies meet the criteria for the SMB paradigm in that they involve manipulations and target behaviors involving morally positive behavior. Other FITD studies utilize nonmoral requests (e.g., requests to participate in market research; Hansen & Robinson, 1980). As the behaviors involved in these requests are not morally salient, they do not meet the criteria for the SMB paradigm and so are not considered evidence for positive consistency as we define it. We limit our review to prosocial-FITD studies.

Prosocial-FITD studies have largely found results that represent positive consistency effects. For example, Foss and Dempsey (1979) found that participants who agreed to help promote a blood drive were subsequently more likely to sign up to donate blood, compared to those who had not been asked to help in the initial promotion. These studies have been cited as evidence for positive consistency within the moral licensing literature (Conway & Peetz, 2012; Cornelissen et al., 2013; Gneezy et al., 2012; Mullen & Monin, 2016). Meta-analyses suggest that behavioral consistency is reliably found both in the FITD literature in general (Beaman et al., 1983; Burger, 1999; Fern et al., 1986; Pascual & Guéguen, 2005) and in the prosocial-FITD literature, in particular (Dillard et al., 1984).

Other Evidence for Positive Consistency. Evidence for positive consistency has also been found outside the prosocial-FITD literature. For example, Aquino et al. (2009) found that participants who wrote about having moral traits engaged in more prosocial behavior than participants who wrote about having neutral traits. It should be noted that this is the same manipulation used by Sachdeva et al. (2009), who found this manipulation resulted in moral licensing, and Blanken et al. (2014), who found no effect on subsequent behavior. Experimental research from the environmental spillover literature has also found support for positive consistency effects. For example, Truelove et al. (2016) found that participants who were induced to recycle subsequently reported higher support for the establishment of a campus green fund, and Geng et al. (2016) found participants induced to choose environmentally friendly products subsequently reported higher pro-environmental intentions compared to participants induced to choose conventional products. This research shows that positive consistency is not constrained to studies using the prosocial-FITD version of the SMB paradigm, and so the specific methodological differences between prosocial-FITD studies and other variations of the SMB paradigm may not be sufficient to explain when moral licensing or consistency is likely to be observed.

Literature on Moderators of Moral Licensing and Consistency

There are a range of individual studies that have explored moderators of moral licensing and positive consistency. These studies have found that factors such as ethical mindset (Cornelissen et al., 2013), focusing on goal progress versus goal commitment (Susewind & Hoelzl, 2014), and whether the manipulation (Kristofferson et al., 2014) or target behavior (Greene & Low, 2014) are performed publicly or privately, moderate whether moral licensing will occur or not (for a comprehensive review of moderators of moral licensing, see Mullen & Monin, 2016). The literature examining moderators of moral licensing consists of a hodgepodge of studies that each focus on distinct moderator variables. No single empirical study has provided an overarching theoretical explanation that encapsulates all potential moderating factors explored in all other individual studies. Effron and Conway (2015) have proposed a theoretical explanation that they argue may integrate the different moderators identified in the literature and provide insight into the broader mechanisms that differentiate the conditions under which each effect may be found.

Theoretical Accounts of Positive SMB Effects and Theoretical Moderators

Moral licensing and positive consistency have been explained using different theoretical models (moral credits, moral credentials, and self-perception theory [SPT]). The only attempt to explain the existence of both moral licensing and positive consistency effects with a single theoretical model was proposed by Effron and Conway (2015). Prior to this work, research on moderators of positive SMB effects has drawn on disparate theoretical backgrounds to make specific predictions about specific manipulation conditions. For example, researchers have examined moderating conditions such as recalling recent versus distant behavior (Conway & Peetz, 2012), moral rules versus the consequences of actions (Cornelissen et al., 2013), commitment to or progress toward goals (Susewind & Hoelzl, 2014), performing behavior in public versus private (Kristofferson et al., 2014), and making a subsequent transgression easier or harder to rationalize (Brown et al., 2011), to name a few (see Mullen & Monin, 2016, for a review of moderators of moral licensing). In each case, researchers have aimed to identify some moderating conditions but have not necessarily tried to apply their findings to explain all variability in SMB effects observed in the literature. In the following section, we briefly outline each of these theories and identify the conditions under which they predict particular SMB effects should and should not occur. Theoretical predictions about the conditions under which each effect may be observed can be tested by coding the methodological characteristics of the studies included in a meta-analysis according to these conditions (or proxies of these conditions) and observing whether effect sizes vary systematically between them. A summary of these moderators and associated predictions is presented in Table 1.

Table 1
Summary of Expected Patterns of Results for Moderators Derived by Different Theoretical Explanations

	Moderator								
	Target behavior valence (proxy for ambiguity)		Domain-consistency		Manipulation behavior type (proxy for effort)			Manipulation focus	
Theoretical explanation	Positive	Negative	Domain-consistent	Domain-inconsistent	Actual behavior	Recalled behavior	Hypothetical behavior	Behavior-focused	Identity-focused
Moral credits	ML	ML (if domain-inconsistent) Null (if domain-consistent)	ML (if target is positive) Null (if target behavior is negative)	ML	X	X	X	X	X
Moral credentials	ML (if domain-consistent) Null (if domain-inconsistent)	Null	ML	Null	X	X	X	X	X
SPT	X	X	PC	Null	PC	X	Null	X	X
Moral identity activation	X	X	X	X	X	X	X	ML	PC

Note. ML = moral licensing; X = no prediction; Null = no difference between moral manipulation condition and control; SPT = self-perception theory; PC = positive consistency.

Moral Credits: An Account of Moral Licensing

Moral credits are the earliest proposed mechanism of moral balancing (Nisan, 1990) and have been applied broadly to explain these effects (Jordan et al., 2011; Merritt et al., 2010; D. T. Miller & Effron, 2010; Sachdeva et al., 2009; Zhong et al., 2010). This model proposed that engaging in moral behavior can positively or negatively impact a person's moral self-image (Merritt et al., 2010; Nisan, 1990, 1991; Zhong et al., 2010), that is, their perceptions of how moral they are (Jordan et al., 2015; Monin & Jordan, 2009). Engaging in a morally positive behavior can earn "moral credit" or a boost to moral self-image. If this results in moral self-image exceeding an equilibrium point at which they feel comfortably moral, a person may be licensed to engage in a negative action without damaging their self-perceptions.

Early theorizing about moral credits did not specify conditions under which moral licensing could occur, but contemporary literature argues that engaging in a positive behavior does not provide a blanket license to engage in negative behavior in all situations. Effron and Monin (2010) and D. T. Miller and Effron (2010) argued that people are unlikely to exhibit moral licensing in conditions in which doing so will make them appear hypocritical. A negative behavior is likely to be seen as hypocritical if it is (a) blatantly immoral (i.e., clearly wrong) and (b) in the same domain as the manipulation (i.e., directly contradicts previous positive behavior). In contrast, transgressions are less likely to be seen as hypocritical if they are ambiguous or are in a different domain to the initial moral behavior. Contemporary theorists argue moral licensing should occur in the latter conditions, but not the former. Although there is no direct evidence to demonstrate this, Effron and Monin (2010) found indirect support for this claim in research measuring participants' judgments of other people's actions. They found that participants judged transgressions as more permissible in conditions in which they would not be seen as hypocritical (ambiguous transgressions or different-domain blatant transgressions), whereas they condemned same-domain blatant transgressions and judged the actors performing them to be hypocrites.

Moderators Based on Moral Credits. The moral credits model predicts that moral licensing should be observed in conditions in which people are less likely to be concerned about appearing hypocritical (Effron & Monin, 2010; D. T. Miller & Effron, 2010). This corresponds to studies in which the target behavior is (a) ambiguous or (b) a blatant transgression but in a different domain to the moral manipulation. In contrast, this model predicts that licensing will not be observed in conditions in which a person is more likely to appear hypocritical. This corresponds to studies in which the target behavior is (a) blatantly immoral and (b) in the same domain as the manipulation.

To test these predictions, we coded studies in terms of target behavior ambiguity and domain-consistency. Only one study (Brown et al., 2011) in the SMB literature has manipulated target behavior ambiguity directly,[1] and no studies have manipulated both ambiguity and domain-consistency. There are many factors that can impact the ambiguity of a transgression (e.g., availability of alternative explanations for behavior, whether behavior is perceived as causing harm, whether the participant values the moral principle the

[1] Effron and Monin (2010) also tested ambiguity, yet, because this study tested participants' *judgments of others' behavior*, rather than their own behavior, it does not constitute a direct test of this prediction.

transgression violates). We did not feel it was appropriate to code ambiguity in accordance with such criteria. Doing so would rely on considerable subjective judgments from coders and require additional information not provided in study reports or not possible to obtain given it requires insight into the participant's perceptions of the situation or their own moral values. This view was also expressed by Blanken et al. (2015) who declined to evaluate ambiguity at all in their meta-analyses for the same reasons.

We chose to use the valence of the target behavior (i.e., positive or negative) as a proxy for ambiguity. This decision was made based on previous research that suggests people perceive moral rules regulating the performance of good deeds as more flexible than those prohibiting bad deeds (Haidt & Baron, 1996; Janoff-Bulman, 2012) and tend to judge sins of omission (i.e., failure to do good) as less wrong and less intentional than sins of commission (i.e., doing something bad; e.g., Bostyn & Roets, 2016; Kordes-de Vaal, 1996; Spranca et al., 1991). That is, sins of omission are perceived more ambiguously, whereas sins of commission tend to be perceived as more blatant. This omission–commission asymmetry in moral judgment is supported by recent meta-analyses (Yeung et al., 2022). We acknowledge that this is only one possible operationalization of ambiguity, but we argue it affords the most objective strategy for coding ambiguity for the present analyses.

To test the predictions of the moral credits model, we also classified effect sizes according to whether they came from studies in which the moral manipulation and target behavior were domain-consistent (e.g., both involve pro-environmental behavior; Truelove et al., 2016) or domain-inconsistent (e.g., a pro-environmental behavior and a cheating task; Mazar & Zhong, 2010). Based on this coding, we then classified studies as fitting moral credits conditions (i.e., positive target behaviors, or domain-inconsistent negative target behaviors) and hypocrisy conditions (i.e., domain-consistent, negative target behaviors). We predicted that moral licensing would be observed in studies that met our operationalization of moral credits conditions, but not in studies that met our operationalization of hypocrisy conditions.

Moral Credentials: An Account of Moral Licensing

The moral credentials model is the second major theory of moral licensing (Effron et al., 2009; Effron & Monin, 2010; Merritt et al., 2010; D. T. Miller & Effron, 2010). According to this account, engaging in an initial positive behavior provides evidence of good moral character (i.e., a person's "moral credentials"), and this can then be used as a basis for reinterpreting subsequent negative behavior so that it is not seen as problematic (Effron & Monin, 2010; Merritt et al., 2010; D. T. Miller & Effron, 2010). The result of this process is that the person no longer perceives the target behavior as immoral, or they do not believe that others will see it as immoral. Thus, they feel they can engage in that behavior.

Moderators Based on Moral Credentials. Two conditions must be met for moral licensing to occur via moral credentials. First, researchers have claimed that the initial positive action must be in the same domain as the target behavior (Effron & Monin, 2010), as the first moral behavior will only allow for the reinterpretation of a subsequent transgression when these two behaviors are conceptually related (i.e., first showing you are not sexist may license subsequent sexual discrimination but will not license stealing money from someone). Second, the target behavior must also be ambiguous (Merritt et al., 2010; D. T. Miller & Effron, 2010). This model proposes that subsequent negative behavior is reinterpreted—this is more easily achieved if the behavior is ambiguous (i.e., it could be immoral, but not necessarily), rather than blatantly immoral (i.e., clearly wrong). Based on the moral credentials model, we predicted that moral licensing would be observed in studies that are both domain-consistent and use ambiguous target behaviors. As previously described, we chose to use the valence of the target behavior as a proxy for ambiguity but acknowledge that this is only one possible operationalization of ambiguity.

Researchers have argued that moral credits and moral credentials are complementary mechanisms of moral licensing, not opposing ones (e.g., D. T. Miller & Effron, 2010; Mullen & Monin, 2016). That is, in situations where moral credentials cannot operate, moral licensing may still occur via moral credits as this mechanism does not require reinterpretation of the target behavior. If this is the case, then it is possible that moral licensing will still be observed outside of the conditions specified here. However, if the moral credentials account is correct, then moral licensing should at least be observed in domain-consistent studies using ambiguous target behaviors.

SPT: An Account of Positive Consistency

SPT (Bem, 1972) is one of the most frequently cited explanations for behavioral consistency effects observed in the FITD literature (e.g., Arnold & Kaiser, 2016; Burger, 1999) and has been used to explain positive consistency effects in the moral balancing literature (Cornelissen et al., 2008; Kristofferson et al., 2014). According to SPT, consistency occurs because people sometimes make inferences about their attitudes, traits, and/or identity from their past behavior. They are then motivated to act in consistent ways to maintain stable self-perceptions. Past behavior may also be used as a heuristic to guide future behavior in uncertain situations. That is, engaging in one action, such as donating to charity, can lead people to infer that these behaviors reflect their character (i.e., that they are the kind of person who gives to charity), and this in turn increases the likelihood that they will behave this way again in the future. This is similar to the process described by the moral credentials account of moral licensing, but it predicts the opposite effect on subsequent behavior. The extent to which one would expect to find positive consistency in an SMB study, according to this theory, depends on the extent to which the initial behavior is diagnostic of the self.

Moderators Based on SPT. According to SPT, moral consistency is likely to occur in conditions that are more conducive to self-perception processes. First, an initial moral behavior that is more effortful or costly is argued to be perceived as more diagnostic of the self and thus more likely to prompt consistency (Burger, 1999; DeJong, 1979; Seligman et al., 1976). In contrast, an initial behavior that requires little effort may not influence self-perceptions, as it is not likely to be viewed as a diagnostic of the self. There is some empirical support for this assertion from FITD studies that have found that larger initial requests are more likely to lead to behavioral consistency than smaller requests (see Burger, 1999, for a meta-analytic review).

In the current meta-analysis, we examined the type of behavior used in the manipulation as a proxy for effort/cost. There are three main types of manipulations in the SMB literature: engaging in an actual behavior (e.g., signing a petition; Kristofferson et al., 2014), recalling past behavior (e.g., Jordan et al., 2011), and hypothetical behavior (e.g., imagining volunteering; Khan & Dhar, 2006). We predicted that manipulation behavior type would moderate effect sizes such that manipulations involving real behavior would lead to positive consistency (as they are the most effortful to perform), whereas those that involve hypothetical actions would not (as they involve little cost or effort). Recall manipulations are relatively effortless for the participant but involve reflection on behaviors that may have been effortful at the time they were performed. As such, this category was included in the analyses for exploratory purposes, but no predictions were made. We acknowledge that this is an imperfect method of categorization (e.g., some performed behaviors still require little effort) and so serves only as a proxy to SPT condition of effort.

A second prediction of SPT is that positive consistency effects should only occur when the moral manipulation and the target behavior are in the same domain (Burger, 1999). If a behavior performed at one point is to be used as a heuristic to guide future decision making, then it must be conceptually related to those future decisions. For example, if past pro-environmental behavior leads a person to infer that they are a “pro-environmentalist,” then it could be expected to influence future pro-environmental decisions, but not unrelated behavior such as stealing. If SPT is correct, then it is predicted that domain-consistency should moderate effect sizes, such that positive consistency should be observed in SMB studies that are domain-consistent, but not studies that are domain-inconsistent.

Activation of Moral Identity: An Explanation of Both Moral Licensing and Positive Consistency

Each of the theories discussed thus far aims to explain the conditions under which specific types of SMB effects occur but do not try to explain when *both* moral licensing and positive consistency should differentially occur. Although many individual articles have been published that apply a range of theoretical backgrounds to predict moderation effects in SMB studies (see Mullen & Monin, 2016, for a review), none of these studies had attempted to provide an overarching theoretical explanation that could explain all variation observed in the literature. To date, only Effron and Conway (2015) have attempted to provide a theoretical explanation integrating all individual moderators of licensing and consistency from past literature. Effron and Conway (2015) proposed that the primary distinction between manipulations that lead to moral licensing and those that lead to positive consistency is whether or not they signal that morality plays an important role in a person’s identity. Specifically, they argue that manipulations that make moral identity salient should lead to positive consistency, whereas those that do not should lead to moral licensing. While this explanation may sound similar to SPT, it differs in that it does not assume people infer identity from past actions. Rather, they will act consistently when moral identity is made salient but may engage in licensing when it is not.

Moderators Based on Activation of Moral Identity. Effron and Conway (2015) argued that manipulations that lead to moral licensing are those that do not make moral identity salient, specifically conditions that involve (a) focusing on the behavior and/or its short-term consequences, (b) focusing on progress toward long-term goals, (c) behaviors that are costless perform, or (d) conditions in which participants are cognitively depleted (and so may be unable to reflect on the connection between behavior and identity). These manipulations are argued to induce focus on the *behavior* and not necessarily on the person performing that behavior. We will refer to these as *behavior-focused* manipulations.

Effron and Conway (2015) suggested that manipulations that make moral identity salient are likely to lead to positive consistency. These are manipulations that involve (a) focusing on abstract moral values or identity, (b) focusing on commitment to long-term goals, (c) costly actions, or (d) conditions in which participants are not cognitively depleted and so are able to reflect on the connection between the behavior and identity. We refer to this second category of manipulations as *identity-focused* manipulations.

In the current meta-analysis, we tested these claims by examining whether the focus of the manipulation used in each study (behavior-focused or identity-focused, coded as per the taxonomy described above) moderated the results. It was predicted that moral licensing would be observed in SMB studies using behavior-focused manipulations and that positive consistency would be observed in SMB studies using identity-focused manipulations.

Exploratory Methodological Moderators

We also examined whether other methodological differences between studies moderated effect sizes. We did not make predictions about these moderators, as they do not relate directly to the conditions outlined by existing theories. Nonetheless, they represent differences in methodological choices between studies that may account for some variability in results.

Target Behavior Type: Behavioral or Hypothetical

SMB studies differ in whether the target behavior involves real (e.g., decisions in economic games; Aquino et al., 2009) or hypothetical behavior (e.g., responses to hypothetical scenarios; Hayley & Zinkiewicz, 2013). As research has shown that participants’ self-reported, hypothetical moral behaviors, and observed moral behavior do not always align (Epley & Dunning, 2000; FeldmanHall et al., 2012; Patil et al., 2011; Teper et al., 2011, 2015), we tested whether effect sizes differed in studies using target behaviors measuring real or hypothetical behavior.

Area of Literature: Prosocial-FITD Versus Non-FITD

The argument that studies from the prosocial-FITD literature are comparable to other SMB studies is based on the rationale that they use the same experimental paradigm as SMB studies outside this literature. Yet there are some methodological differences that could potentially account for differences in results between these studies and SMB studies outside this literature. For example, prosocial-FITD studies always use behavioral manipulations (i.e., participants perform an actual behavior), positive target behaviors (e.g., requests to help, volunteer), and involve an escalation of behavior (i.e., the target behavior involves a larger request than the manipulation). In comparison, SMB studies outside the prosocial-FITD literature tend to use different variations of manipulations (e.g., hypothetical, recall, or real behavior tasks), use a mix of positive and negative target behaviors, and do not necessarily involve an escalation of tasks. As a way of capturing these broad differences in methodology, we tested whether effect sizes differed systematically between prosocial-FITD studies and SMB studies published outside of this literature (i.e., prosocial-FITD vs. non-FITD). We emphasize that FITD does not refer to all FITD studies but only *prosocial*-FITD studies (i.e., those including requests relating to moral actions).

Audience for Manipulation and Target Behavior

In some SMB studies, efforts are made to make moral manipulation and target behavior appear unrelated. This may be by ostensibly presenting them as part of two different studies being conducted for separate research teams, but with a shared data collection session (e.g., Jordan et al., 2011), or by having each task presented by two seemingly unrelated people (e.g., Dolinski, 2012). In essence, this represents an attempt to make the tasks appear as though they are being observed by two distinct audiences. In most studies, the relationship between the manipulation and target behavior is not disclosed, but no effort is made to make them appear as though they are being observed by different audiences.

Existing theories do not make explicit predictions about whether having the same or different audiences for the manipulation and target behavior should impact SMB effects. However, having the same audience may introduce a source of hypocrisy concerns, which the moral credits model suggests may suppress licensing (Effron & Monin, 2010). Evidence of the effect of audience is mixed. On the one hand, Monin and Miller (2001) found that moral licensing occurred both in conditions in which the moral manipulation and target behavior were observed by the same audience and conditions in which they were observed by ostensibly unrelated audiences. On the other hand, Kristofferson et al. (2014) found moral licensing only occurred if both the manipulation and target behavior were performed publicly (i.e., observed by the same audience) but found positive consistency occurred when the manipulation involved behaviors that were performed privately (i.e., not observed by the same audience as the target behavior). Greene and Low (2014) found that moral licensing only occurred when the target behavior (engagement in hypothetical moral transgressions) was described to participants as being private (i.e., not observed by others). We examined whether efforts to make participants believe the moral manipulation and target behavior were observed by unrelated audiences would moderate effect sizes. We did not make specific predictions about how the audience would moderate effect sizes due to both the lack of theoretical predictions regarding this and the inconsistency in the results of previous literature.

Year of Publication

Prosocial-FITD research dates to the 1960s (e.g., Freedman & Fraser, 1966) and typically reports evidence of positive consistency. The concept of moral balancing was not formally discussed until much later (Nisan, 1990) and did not come to prominence until the early 2000s with the publication of Monin and Miller's (2001) research. Due to the disparity in when these effects emerged, we chose to examine whether effect sizes were moderated by publication year. Changes in results over time could reflect changes in social norms relating to moral behavior, changes in research practices (e.g., greater emphasis on larger sample sizes and statistical power), or merely a change in research interest from using the specific variation of SMB design seen in prosocial-FITD literature to more novel variations seen in moral licensing literature. We also examined whether there were relationships between sample size and publication year, and sample size and area of literature, as a strategy for investigating differences in research practices and quality over time.

Method

Transparency and Openness

We did not preregister a protocol for these meta-analyses. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses reporting guidelines for the final report. Additional online materials can be found at https://osf.io/upby2/. These additional online material include a full list of search terms with comparisons to previous meta-analyses (Table S1: https://osf.io/yb589), a list of studies excluded during full-text screening with exclusion reason (Table S2: https://osf.io/yb589), and effect size information (study characteristics, effect size data, moderator coding) for both Meta-Analysis 1 (Table S3: https://osf.io/yb589) and Meta-Analysis 2 (Table S5: https://osf.io/yb589). Data files and the analysis script can also be found via this link (https://osf.io/upby2/).

Literature Search

We conducted a systematic search of the following databases: APA PsycInfo, Proquest, Medline, and Web of Science Core Collection.[2] The search was conducted to identify studies for both Meta-Analysis 1 (positive manipulations) and Meta-Analysis 2 (negative manipulations) concurrently and so it included terms relevant to both. The search was originally conducted in April 2015 and updated in January 2019, November 2020, and July 2022.

SMB studies come from a range of literature that use different terms to describe research designs and effects, as the term SMB paradigm was not coined until relatively recently (Mullen & Monin, 2016). As such, we used multiple search strings to target different areas of literature known to publish SMB effects. The first string targeted moral balancing literature and combined terms for three concepts: (a) morality (e.g., *moral**), (b) observed effects (e.g., *licens**), and (c) outcome measures of moral and immoral behavior (e.g., *help**). The second search string targeted the sequential request and guilt and compliance literature and included terms relating to two concepts: (a) effect terms relating to FITD, guilt and compliance, and door-in-the-face (DITF) literature (e.g., *FITD*) and (b) moral behavior (e.g., *prosocial*). The third search targeted the environmental spillover literature and included terms relating to two concepts (a) spillover (e.g., *spillover* OR *spill-over*) and (b) environmental behavior (e.g., *pro-environment**). All search terms can be found in the additional online material (Table S1: https://osf.io/yb589). Although these terms are not exhaustive, they were adequate to capture a broad representation of SMB studies. Additional online material (Table S1: https://osf.io/yb589) also provides a comparison of search terms used in previously published

[2] At the time of the searches, the institutional subscription to Web of Science covered Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index, Conference Proceedings Citation Index—Science, Conference Proceedings Citation Index—Social Science and Humanities, Book Citation Index—Science, Book Citation Index—Social Sciences and Humanities, Emerging Sources Citation Index, Current Chemical Reactions, and Index Chemicus.

meta-analyses on related topics (e.g., moral licensing, FITD). We also searched reference lists of included studies and review articles and meta-analysis on related topics (e.g., Blanken et al., 2014; Geiger et al., 2021).

In addition to the database search, calls for unpublished data were distributed through the list servs and online forums of the Australasian Society for Social Psychology, the Society for Personality and Social Psychology, and the European Association for Social Psychology. We also contacted prominent authors in the field directly to enquire about unpublished data.

Study Selection

Screening for both Meta-Analyses 1 and 2 was conducted concurrently. First, the titles and abstracts of all records located through the search strategy were screened for relevance. Those selected for full-text screening were screened by the first author, and a subset of 110 records were double-screened by other members of the research team. Any disagreements were resolved through discussion with a third screener.

The inclusion criteria for Meta-Analysis 1 were as follows:

1. Participants had to be individuals. Studies measuring household behavior (e.g., household resource use) or decisions by organizations were excluded.

2. Studies had to use a between-groups, experimental design using the SMB paradigm, with random allocation to experimental conditions (correlational studies or studies in which participants were grouped based on initial behavior, rather than it being manipulated, were excluded).

3. Studies had to include a positive moral manipulation based on actual, recalled, or imagined behavior. We also included SMB studies using trait-based manipulations (e.g., writing about having positive moral traits; Sachdeva et al., 2009). These manipulations are argued to directly impact participants' moral self-image and have been used in studies that have found both moral licensing (Sachdeva et al., 2009) and positive consistency (Aquino et al., 2009). Studies that used manipulations based on nonmoral behavior or other people's behavior were excluded.

4. Studies had to include a control condition in which participants did not perform, imagine, or recall their own moral behavior. We included studies in which participants in the control condition either (a) performed, recalled, or imagined a morally neutral behavior, (b) observed, recalled, or imagined someone else's behavior, or (c) were passive controls (performed no manipulation tasks). We excluded studies that compared positive manipulation conditions to negative manipulation conditions (i.e., no control group) to avoid ambiguity in interpreting results (for further discussion on the need for neutral control conditions in SMB research, see Mullen & Monin, 2016).

5. Studies had to include a morally positive or negative target behavior that was measured in the same experimental session as the moral manipulation. Studies in which the moral manipulation and target behavior were measured in separate sessions, or longitudinal studies tracking behavior over time were excluded to avoid the possibility that intervening events may confound the results. This reasoning is grounded in assumptions of the moral credits model, which proposes each instance of engaging in a morally salient behavior has the potential to change moral self-image and thus future moral behavior. If there is a substantial temporal gap between the manipulation and target behavior, there is an opportunity for participants to engage in additional morally salient actions that could alter their moral self-image and contaminate the results. This means it would not be possible to determine if the participant's behavior is due to the moral manipulation or due to other actions that participants have engaged in during the intervening time.

6. Reports had to be written in English or have verified English translations available. This was to avoid any errors resulting from incorrect translation by the research team. Only two reports were excluded for this reason.

Data Extraction

The first author completed data extraction for all included studies. Approximately 41% of the included studies ($k = 91$) were double-extracted by the third or fourth author. Discrepancies in data extraction were minimal (two effect sizes differed between coders due to an error in extracting cell sizes; these discrepancies were resolved by referring to the study reports). Interrater agreement was high for coding of moderators (see Table 2). Due to the high level of agreement between extractors and due to practical constraints, it was decided that the remaining effect sizes would be extracted by the first author only.

Study Characteristics. For each study, we recorded a description of the moral manipulation, control condition, and target behavior (see additional online material, Table S2: https://osf.io/yb589). If multiple target behaviors were included in a study, we extracted data from the first measure presented to participants when the presentation occurred in a fixed order. If the order of presentation was randomized ($k = 4$ effect sizes), we averaged these together into a single effect size.

Moderator Coding. Each moderator category was coded as described below.

Target Behavior Valence. Studies were coded to identify whether they included a positive (e.g., charitable donations) or negative target behavior (e.g., cheating).

Domain-Consistency. Studies were coded as domain-consistent if the manipulation and target fit into the same broad conceptual category (e.g., pro-environmental behavior; racial prejudice; sharing). They were coded as domain-inconsistent if the manipulation and target did not fit the same broad category. Where the domain of the manipulation was not possible to identify (e.g., recall manipulations that just ask participants to recall acting morally; Stellar & Willer, 2014), the studies were coded as uncategorized and omitted from the analysis of this moderator.

Manipulation Behavior Type. Manipulations were classified into one of three categories: (a) performed behavior (e.g., signing a petition; Kristofferson et al., 2014), (b) recalled behavior, or (c) hypothetical behavior, such as responding to hypothetical choices (e.g., Berger et al., 2020), agreeing to perform an action in the future (Cascio & Plant, 2015), or writing about themselves using moral

Table 2
Descriptive Statistics for Positive Sequential Moral Behavior Studies, Meta-Analysis 1

Moderator	*k*	Sample size *M*	*SD*	*Mdn*	Range	Publication year *M*	*SD*	*Mdn*	Range
All studies	217	143.97	248.02	84.00	16–3,286	2011.63	8.27	2013	1972–2022
Target behavior valence (κ = .98)									
Negative	67	159.97	402.99	80.00	31–3,286	2013.54	4.61	2012	2001–2021
Positive	150	136.83	130.65	84.50	16–708	2011.23	9.45	2014	1972–2022
Domain-consistency (κ = .82)									
Inconsistent	47	158.66	154.38	86.00	29–687	2012.40	7.79	2014	1981–2021
Consistent	107	156.30	326.96	86.00	31–3,286	2010.05	10.20	2013	1972–2022
Manipulation behavior type (κ = .79)									
Hypothetical	93	159.11	351.43	84.00	16–3,286	2013.00	4.44	2015	2001–2020
Recall	45	139.67	148.05	76.00	29–708	2013.15	11.79	2013	2008–2022
Behavioral	75	128.20	110.40	84.00	29–686	2008.12	12.13	2012	1972–2021
Manipulation focus (κ = .83)									
Identity-focused	55	110.16	119.11	67.00	16–567	2012.2	3.75	2013	2001–2020
Behavior-focused	155	155.25	281.72	86.00	29–3,286	2011.30	9.49	2014	1972–2022
Target behavior type (κ = .80)									
Hypothetical	105	163.30	337.39	82.00	29–3,286	2011.50	8.09	2012	1975–2022
Behavior	112	125.86	111.17	88.00	16–686	2011.76	8.47	2014	1972–2021
Area of literature (κ = .92)									
Non-FITD	176	157.42	272.80	85.00	16–3,286	2013.77	4.21	2014	2001–2022
Prosocial-FITD	37	89.78	49.90	74.00	29–229	2001.29	13.85	2006	1972–2018
Audience (κ = .75)									
One audience	138	170.61	301.72	93.00	18–3,286	2011.87	8.71	2014	1972–2022
Two audiences	28	74.39	39.72	64.00	16–157	2008.07	11.67	2012	1977–2021
Publication status (κ = 1.00)									
Unpublished	58	116.93	100.45	80.50	29–447	2013.25	2.54	2014	2010–2020
Published	159	153.84	282.99	84.00	16–3,286	2010.80	9.42	2013	1972–2022

Note. Outliers are omitted from Meta-Analysis 1. FITD = foot-in-the-door.

trait words (Sachdeva et al., 2009). Manipulations that included elements of multiple categories (e.g., recalling previous behavior and writing about having moral traits; Blanken et al., 2012) were omitted from moderation analyses.

Manipulation Focus. Manipulations were coded dichotomously as either behavior-focused or identity-focused according to criteria identified by Effron and Conway (2015). Specifically, behavior-focused manipulations were those that (a) were accompanied by instructions or primes to think about the behavior concretely (e.g., to focus on immediate consequences; Cornelissen et al., 2013), (b) were accompanied by instructions or primes to think about progress toward goals (Geng et al., 2016; Susewind & Hoelzl, 2014), or (c) involved simply performing or recalling a behavior without additional instructions to reflect on these actions in a particular way (e.g., Jordan et al., 2011).

Identity-focused manipulations were those that involved instructions or priming conditions aimed to induce participants to (a) think about the behavior abstractly (e.g., focus on moral values or rules or temporally distant behavior; Conway & Peetz, 2012; Cornelissen et al., 2013), (b) think about long-term traits (e.g., write about their traits; Sachdeva et al., 2009), (c) think about commitment to, or how to achieve long-term goals (e.g., Susewind & Hoelzl, 2014). This category also included manipulations that involved providing participants with feedback that ostensibly reflected their identity or character, for example, telling them that they are "helpful" or "environmental" people (Cornelissen et al., 2007; Guadagno et al., 2001).

Where manipulations included elements of both behavior- and identity-focused manipulations they were left unclassified and omitted from the analysis of this moderator. For example, Mann and Kawakami (2011) provided false feedback that participants were progressing toward a goal (behavior-focused) of becoming less prejudiced people (identity-focused).

Target Behavior Type. Studies were coded according to whether the target behavior involved the performance of actual behavior (e.g., donating real money to charity) or hypothetical behavior (e.g., responses to hypothetical dilemmas). Outcomes that involved asking the participant to agree to perform a behavior in the future (e.g., sign up to volunteer; Kristofferson et al., 2014) were coded as real behavior when they were presented to participants as genuine requests for future action rather than as hypothetical requests (i.e., participants believed they would actually engage in the behavior in the future; Goldman et al., 1982) and as hypothetical when they were framed as general intentions (e.g., future prosocial intentions; Jordan et al., 2011).

Area of Literature. Studies were coded according to whether they came from the prosocial-FITD literature or whether they were published outside of this literature (non-FITD). Categorization was based on the authors' description of their research (i.e., whether they identified it as an FITD study or not).

Audience for Manipulation and Target Behavior. Studies were coded as having separate audiences for the manipulation and target behavior if (a) participants were explicitly told that the manipulation and target behavior were part of separate, unrelated studies or (b) the manipulation and target behavior were presented by different, ostensibly unrelated people. Studies in which no attempt to create the impression of separate audiences was made were coded as having one audience. Studies in which there was insufficient information to determine whether an attempt to portray separate audiences had been made or where the number of audiences was unclear were omitted from this moderation analyses.

Year. Year was coded as per the official publication date on journal websites. For unpublished studies, year was coded as the year data were collected or the year unpublished articles were prepared.

Meta-Analytic Procedure

Effect Size Data and Calculations for Individual Studies. The effect size metric used in the current analyses is Hedges' *g* (a standardized mean difference that has been corrected for sampling bias; Hedges, 1981), reported with 95% confidence intervals. All effect sizes were coded such that positive values indicated a moral licensing effect (i.e., an increase in subsequent negative behavior after a manipulation, relative to the control) and negative values indicated a positive consistency effect (i.e., a decrease in subsequent negative behavior after a manipulation, relative to the control). We interpret the magnitude effects by comparing them to percentiles of effects typically reported in social psychology (Lovakov & Agadullina, 2021), as SMB studies are typically published in social psychology or adjacent fields (e.g., organizational behavior).

For continuous target behaviors (e.g., amount of cheating; Mazar & Zhong, 2010), effect sizes were calculated from cell sizes, means and variance (*SD*s or *SE*s) wherever possible, or inferential statistics (e.g., *t* tests) when descriptive statistics were not available. For dichotomous target behaviors (e.g., agree to volunteer or not; Kristofferson et al., 2014), we extracted cell sizes and frequencies where available, or chi-square statistics where frequencies were unavailable. These data were entered into the program Comprehensive Meta-Analysis (Borenstein et al., 2014) to calculate Hedges' *g* values and effect size variance. Where insufficient statistical data were reported in study records, we contacted study authors to request missing information. If no response was received, then data were estimated, when possible (e.g., dividing the overall sample size by the number of conditions), or the study was excluded if estimation was not possible. We include notes indicating which effect sizes required estimation in the additional online material (Table S3: https://osf.io/yb589).

Point Estimate Calculations. The individual effect sizes extracted from included studies were aggregated into point estimates (i.e., average effect sizes) using a random-effects model with robust variance estimation (RVE; Hedges et al., 2010). As with individual effect sizes, the point estimate used was Hedges' *g* with 95% confidence intervals. Effect sizes were coded such that a positive point estimate indicates moral licensing (i.e., an increase in negative behavior), whereas a negative point estimate indicates positive consistency (i.e., a decrease in negative behavior). Heterogeneity between effect sizes was quantified with τ, which represents the estimated standard deviation of the aggregated effect size estimate (Borenstein et al., 2009).

We selected RVE methods to combine effect sizes to account for two types of dependency present in the data: correlated effects and hierarchical relationships between effect sizes and studies. RVE accounts for these kinds of dependency by adjusting effect size weights, even when the degree and structure of covariance between effect sizes are unknown (Hedges et al., 2010). First, we used a correlated effects model (CEM; Fisher & Tipton, 2015) to account for dependency due to correlated effects. A small but notable number of effect sizes ($k = 19$) were calculated from studies that included multiple moral manipulation conditions, but a single control group (e.g., Experimental Group 1 vs. Control, Experimental Group 2 vs. Control). In these cases, effect sizes were calculated for each moral manipulation condition in the study. As these effect sizes are calculated using the same control group, they are correlated and thus nonindependent. CEM models account for this dependency by calculating weights based on sample clusters. Effect sizes derived from shared control groups are grouped into the same sample clusters, and effect sizes derived from independent samples are counted as individual clusters.

Second, we used a hierarchical effects model (HEM; Fisher & Tipton, 2015) to account for dependency due to hierarchical relationships between effect sizes and studies. There were a large number of effect sizes calculated from independent groups of participants that were nested within studies. These studies included multiple moral manipulation conditions and multiple control groups (e.g., Experimental Group 1 vs. Control Group 1; Experimental Group 2 vs. Control Group 2). In cases like this, even though effect sizes are based on the different groups of participants, there may still be dependency between them as they emerged from the same studies. HEM estimates account for this kind of dependency by calculating weights based on study clusters. Effect sizes derived from the same studies are grouped into the same clusters. The results of both CEM and HEM RVE models are presented throughout the results. RVE analyses were conducted using the *robumeta* package in R (Fisher & Tipton, 2015).

Outliers. We ran influence diagnostics using the *metafor* package in R to identify outliers and quantify their influence on the data before combining effect sizes (Viechtbauer, 2015). Cases with studentized residuals >±1.96 were identified as outliers. Outliers with Difference in Fits (DFFITs) values >±0.20 (i.e., $3 \times \sqrt{\frac{p}{k-p}}$) where *p* is equal to the number of estimated coefficients and *k* is equal to the number of observations) were considered influential (Viechtbauer & Cheung, 2010). Influential cases were removed prior to analyses.

Moderation Analyses. Moderators were coded categorically (see descriptions above) and dummy-coded as 0 and 1 (see Table 3). Manipulation behavior type consisted of three levels (hypothetical, recalled, or actual behavior), so two dummy-coded comparisons were created using the hypothetical manipulations as the reference category (coded as 0). Year was entered as a continuous moderator.

We tested whether effect sizes differed systematically by moderator group using metaregression with RVE using both CEM and HEM weights. We began by testing each moderator category separately. We then entered significant moderators simultaneously into a metaregression to identify whether each moderator uniquely accounted for differences in effect sizes, or whether the results of the initial simple metaregressions were partially driven by overlap between moderator categories.

File-Drawer Bias. File-drawer bias refers to the issue of missing data from unreported studies that are typically more likely to find smaller effects and nonsignificant results (Rosenthal, 1979). Omitting these studies from meta-analyses can result in overestimation of the reliability and magnitude of effect sizes. We did receive some reports of unpublished results from authors in the field (~26% of effect sizes included in Meta-Analysis 1 were unpublished. The unpublished

Table 3
Meta-Analysis 1: Average Effect Size and Heterogeneity Estimates for Full Sample and Moderator Categories

Moderator	*k*	Correlated effects RVE model					Hierarchical weights RVE model				
		SmC	*g*	95% CI	τ	Q_B [95% CI]	StC	*g*	95% CI	τ	Q_B [95% CI]
Target behavior valence						**−0.32 [−0.44, −0.21]**					**−0.32 [−0.43, −0.20]**
Negative (0)	67	65	**0.25**	**[0.16, 0.44]**	0.23		52	**0.24**	**[0.15, 0.33]**	0.23	
Positive (1)	150	141	**−0.07**	**[−0.14, 0.004]**	0.32		110	**−0.07**	**[−0.15, −0.003]**	0.32	
Domain-consistency						**−0.23 [−0.39, −0.07]**					**−0.23 [−0.38, −0.08]**
Inconsistent (0)	47	46	**0.17**	**[0.04, 0.30]**	0.31		37	**0.16**	**[0.04, 0.29]**	0.19	
Consistent (1)	107	100	−0.06	[−0.15, 0.03]	0.33		78	−0.07	[−0.16, 0.02]	0.32	
Manipulation behavior type											
Contrast 1: Hypothetical (0) versus recall (1)						**−0.16 [−029, −0.02]**					**−0.17 [−0.31, −0.04]**
Contrast 2: Hypothetical (0) versus behavior (1)						**−0.32 [−0.46, −0.18]**					**−0.33 [−0.47, −0.20]**
Hypothetical	93	92	**0.17**	**[0.08, 0.26]**	0.27		77	**0.17**	**[0.08, 0.26]**	0.27	
Recall	45	42	−0.01	[−0.10, 0.11]	0.25		34	−0.003	[−0.11, 0.10]	0.25	
Behavioral	75	68	−0.16	**[−0.27, −0.05]**	0.37		54	**−0.17**	**[−0.28, −0.07]**	0.32	
Manipulation focus						0.10 [−0.04, 0.23]					0.11 [−0.03 0.24]
Identity-focused (0)	55	55	−0.06	[−0.17, 0.04]	0.27		48	−0.06	[−0.17, 0.04]	0.27	
Behavior-focused (1)	155	150	0.04	[−0.03, 0.11]	0.32		122	0.04	[−0.03, 0.11]	0.31	
Target behavior type						−0.08 [−0.20, 0.04]					−0.09 [−0.21, 0.03]
Hypothetical (0)	105	104	0.07	[−0.006, 0.16]	0.28		75	0.07	[−0.01, 0.15]	0.28	
Behavior (1)	112	103	0.01	[−0.09, 0.08]	0.34		88	−0.02	[−0.11, 0.07]	0.33	
Area of literature						**−0.54 [−0.71, −0.38]**					**−0.53 [−0.64, −0.42]**
Non-FITD (0)	176	172	**0.11**	**[0.05, 0.17]**	0.27		133	**0.11**	**[0.05, 0.17]**	0.27	
Prosocial-FITD (1)	37	32	**−0.44**	**[−0.59, −0.29]**	0.35		26	**−0.43**	**[−0.54, −0.33]**	0.40	
Audience						−0.09 [−0.32, 0.11]					−0.10 [−0.34, 0.14]
One audience (0)	138	132	0.04	[−0.04, 0.11]	0.31		103	0.03	[−0.04, 0.10]	0.30	
Two audiences (1)	28	28	−0.09	[−0.31, 0.14]	0.47		21	−0.09	[−0.32, 0.14]	0.30	
Year						**0.01 [0.001, 0.01]**					**0.01 [0.001, 0.02]**

Note. Significant effect sizes and regression coefficients are bolded. RVE = robust variance estimation; *k* = total effect sizes; SmC = sample clusters; *g* = Hedges' *g* average effect size estimates for studies within moderator subgroup (positive values represent a moral licensing effect; negative values represent a positive consistency effect); CI = confidence interval; Q_B = regression coefficients for tests of moderators; StC = study clusters; FITD = foot-in-the-door.

studies came from a small number of researchers and were exclusively for non-FITD studies. As such, it is likely these represent only a portion of unreported SMB studies that have been conducted. Thus, we were interested in assessing whether our analyses were impacted by file-drawer bias. We emphasize here that we were not interested in assessing if *publication* bias was present in the published literature, which would involve only analyzing published results. Instead, we were interested in evaluating if *our analysis*, which incorporated both published and unpublished effects, may be biased because there are still unreported studies that we were unable to locate and include. To that end, we conducted bias analyses on the full data set, including both published and unpublished effects.

We used contour-enhanced funnel plots to examine if bias may be present in the data (e.g., Palmer et al., 2008). These graphs plot effect size against standard errors and then overlay contours that represent regions of conventional statistical significance. Unreported studies are typically more likely to be nonsignificant, so if studies appear to be absent from regions of nonsignificance, this suggests that significant results are overrepresented in the effect sizes included in the analyses, and nonsignificant results (that would be expected even if an effect is real) are systematically missing. Contour-enhanced plots were produced using the *metafor* package in R (Viechtbauer, 2015).

Next, we conducted precision-effect test and precision-effect estimate with standard errors (PET-PEESE; Stanley & Doucouliagos, 2014) and a three-parameter selection model (3-PSM; Hedges & Vevea, 1996) to statistically identify if estimates are likely to be impacted by file-drawer bias and calculated revised effect size estimates that attempt to correct for this bias. PET-PEESE uses a weighted regression model (linear for PET, quadratic for PEESE) that predicts effect sizes from their standard errors. If the slope of the regression model is significant, this indicates the data are likely impacted by file-drawer bias. If this is the case, the model intercept provides an effect size corrected for file-drawer bias. When selecting whether to use PET or PEESE estimates, we adhered to conditional logic recommended by Stanley and Doucouliagos (2014): When PET produced a nonsignificant effect size estimate, we retained these estimates; where PET produced a significant effect size estimate, we used PEESE estimates instead. PET-PEESE was conducted using the *metafor* package in R (Viechtbauer, 2015).

We used a 3-PSM (Hedges & Vevea, 1996) using the *weightr* package in R (Coburn & Vevea, 2017). 3-PSM models file-drawer bias by estimating the likelihood of significant results being present (selection parameter) and then estimating the mean effect size and heterogeneity of effect sizes (data parameters) when file-drawer bias is present. If the fit of the model that assumes file-drawer bias is significantly better than the fit of a model that assumes no bias, then this is taken as evidence of bias and the modeled effect size estimate is interpreted as a corrected "true" effect size estimate.

Results

Study Characteristics

We extracted $k = 220$ effect sizes for meta-analysis. Ten effect sizes were identified as outliers based on their studentized residual values. Of these, three effect sizes had DFFITs values that suggested they were exerting influence on the data and so were excluded from the analyses prior to combining effect sizes. These effects are identified in the additional online material (see Table S3: https://osf.io/yb589). The final analyses included $k = 217$ effect sizes from 161 study clusters and 206 sample clusters. The total sample size was $N = 31{,}242$.

Table 2 presents descriptive statistics across the full sample of included effect sizes and within moderator categories, as well as Cohen's κ values for interrater agreement for moderator coding. The most frequently used target behaviors were positive ($n = 150$, compared to $n = 67$ using negative target behaviors) and in the same domain as the moral manipulation ($n = 107$, compared to $n = 47$ domain-inconsistent target behaviors). There were a roughly equal number of effect sizes from studies measuring hypothetical ($n = 105$) and actual behavior target behaviors ($n = 112$). Most moral manipulations were classified as hypothetical ($n = 93$), followed by behavioral ($n = 75$), and then recall ($n = 45$), and more manipulations were classified as behavior-focused ($n = 155$), compared to identity-focused ($n = 55$). Most effect sizes came from studies in the non-FITD literature ($n = 176$, vs. $n = 37$ prosocial-FITD), that used one audience ($n = 138$, vs. $n = 28$ two audiences), and were extracted from published journal articles ($n = 159$, vs. $n = 58$ unpublished).

Included studies ranged in sample size from $n = 16$–3,286, and publication dates ranged from 1972 to 2022. Average sample size was skewed by the inclusion of Ebersole et al. (2016), which had an unusually large sample compared to other studies ($n = 3{,}286$). Median sample sizes were similar across moderator categories, but lower for effect sizes using recall manipulations, identity-focused manipulations, published in the prosocial-FITD literature, or involving two audiences. Average publication year was also similar across moderator categories, but lower and more variables for studies published in the prosocial-FITD literature or using two audiences.

Search and Screening

The results of the search and screening processes are presented in Figure 2. A complete list of excluded studies (with reasons) is provided in the additional online material (Table S2: https://osf.io/yb589). A summary of individual effect sizes included in Meta-Analysis 1, including descriptions of manipulations, outcomes, and moderator coding effect size data, is provided in the additional online material (Table S3: https://osf.io/yb589).

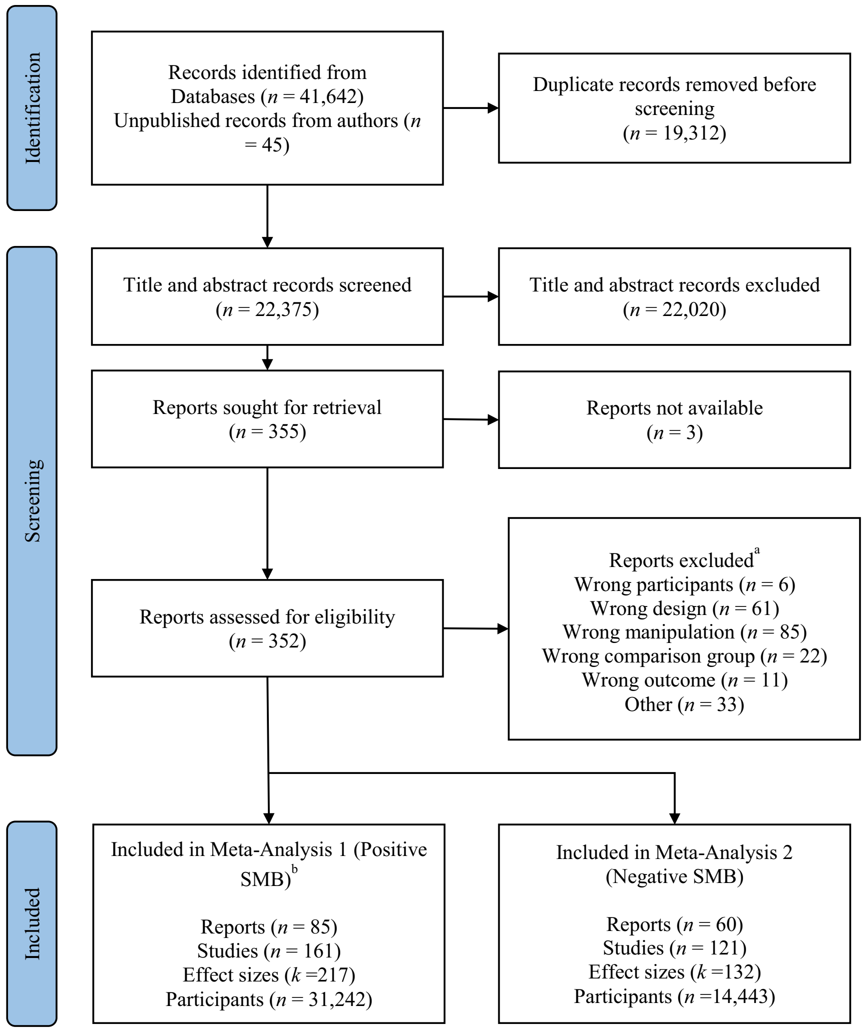
Preliminary Analyses

Ten effect sizes were identified as outliers based on their studentized residual values. Of these, three effect sizes had DFFITs values that suggested they were exerting influence on the data and so were excluded from the analyses prior to combining effect sizes. These effects are identified in the additional online material (see Table S3: https://osf.io/yb589).

The final analyses included $k = 217$ effect sizes from 161 study clusters and 206 sample clusters. Both the CEM and HEM estimates indicate that the average effect size for all studies included in Meta-Analysis 1 was not significantly different from zero (CEM Hedges' $g = 0.04$, 95% CI [−0.02, 0.10], $p = .238$, $\tau = 0.31$; HEM Hedges' $g = 0.03$, 95% CI [−0.04, 0.09], $p = .423$, $\tau = 0.32$). We report these results for completeness but note that they may not be particularly informative considering they consist of an amalgamation of effects from different literatures. Interpreting a main effect in this context

Figure 2

Flow of Reports and Studies Into the Meta-Analyses



Note. SMB = sequential moral behavior. See the online article for the color version of this figure. [a] See additional online material (https://osf.io/upby2/) for detailed reasons. [b] Excluding three outliers.

is problematic for several reasons. First, each literature typically finds results in opposite directions, and so these may cancel each other out when aggregating across them. Second, if an overall effect size did favor a particular type of SMB effect, it would be difficult to differentiate between whether this would reflect (a) that SMB studies do tend to be more likely to find effects in a particular direction, regardless of whether they are labeled as FITD studies or not, (b) that there may simply be more studies in one literature compared to the other, (c) each literature is differentially impacted by file-drawer bias, or (d) some combination of all of these issues. Given the difficulty in interpreting such a main effect, we focus our analysis instead on examining potential moderators of SMB effects.

Moderator Analyses

Methodological Moderators. We first tested each moderator separately using a series of metaregressions (see Table 3). Where Q_B is significant, this indicates a significant metaregression slope—this in turn indicates that effect sizes differ significantly between the moderator groups (for categorical moderators) or that effect sizes vary as a function of continuous moderators. Estimates of average effect sizes (g) for studies within each moderator category are also provided in Table 3.

As Table 3 shows, there were significant moderator effects for target behavior valence, domain-consistency, manipulation behavior type, and area of the literature. Specifically, we found average effect sizes representing significant moral licensing effects for studies using negative target behaviors (equivalent to the 40–45th percentile of effect sizes in social psychology; Lovakov & Agadullina, 2021) and in studies that were domain-inconsistent, that used hypothetical manipulations, or were published outside of the prosocial-FITD literature (all estimates ranged between the 20–30th percentile social psychology effects). In contrast, we found average effect sizes representing significant positive consistency effects for studies

Table 4
Metaregression Coefficients and Moderator Dummy Coding

	Dummy coding		CEM RVE model		HEM RVE model	
Moderator	(0)	(1)	β	95% CI	β	95% CI
Model 1 (k = 133)						
Intercept			0.36	[0.23, 0.50]	0.36	[0.23, 0.50]
TBV	Negative	Positive	**−0.26**	**[−0.44, −0.08]**	**−0.26**	**[−0.42, −0.10]**
DC	Inconsistent	Consistent	−0.03	[−0.18, 0.13]	−0.03	[−0.16, 0.12]
MBT	Hypothetical	Behavior	−0.13	[−0.31, 0.05]	−0.13	[−0.30, 0.12]
LIT	Non-FITD	Prosocial-FITD	**−0.37**	**[−0.60, −0.17]**	**−0.38**	**[−0.54, −0.21]**
Model 2 (k = 214)						
Intercept			**0.25**	**[0.1816, 0.34]**	**0.24**	**[0.15, 0.33]**
TBV	Negative	Positive	**−0.22**	**[−0.34, −0.11]**	**−0.21**	**[−0.32, −0.09]**
LIT	Non-FITD	Prosocial-FITD	**−0.46**	**[−0.63, −0.29]**	**−0.45**	**[−0.57, −0.33]**

Note. Bolded values are significant at the p < .05 level. See the text for an explanation of β coefficients. CEM = correlated effects model; RVE = robust variance estimation; HEM = hierarchical effects model; CI = confidence interval; TBV = target behavior valence; DC = domain-consistency; MBT = manipulation behavior type; LIT = area of literature; FITD = foot-in-the-door.

using positive target behaviors (equivalent to approximately the 15th percentile) and prosocial-FITD studies (equivalent to approximately the 55–60th percentile).

Unique Effects of Moderators. The results of the subgroup analysis may be confounded by overlap between moderator categories. To disentangle this overlap, we entered each of the significant methodological moderators (target behavior valence, domain-consistency, manipulation behavior type, area of literature) simultaneously into a metaregression (see Table 4 for moderator subgroup dummy codes and regression coefficients).

The results of the metaregression indicate that, after controlling for overlap between moderators, effect sizes vary as a function of target behavior valence and area of literature, but not domain-consistency or manipulation behavior type. We dropped the nonsignificant moderators and reran the regression to obtain regression coefficients that average across these conditions (see Model 2, Table 4).[3] As Table 5 shows, the average effect size for studies that use negative manipulations (k = 67; all non-FITD studies) was a moral licensing effect equivalent in magnitude to the 40th–45th percentile for social psychology effects (Lovakov & Agadullina, 2021). The average effect size for studies using positive target behaviors differed by area of literature. Specifically, for prosocial-FITD studies (k = 37), the average effect size represented a positive consistency effect (55th–60th percentile), whereas for non-FITD studies using positive target behaviors (k = 109), the average effect size was close to zero.

Moral Credits Conditions. To test predictions of the moral credits account, we evaluated average effect sizes for studies that met our operationalization of the conditions for hypocrisy (i.e., studies that were both domain-consistent and used negative target behaviors [our proxy for blatant transgressions]) and studies that met our operationalization for the conditions for moral credits (i.e., domain-inconsistent studies and domain-consistent studies that also used positive target behaviors [our proxy for ambiguity]). Contrary to predictions, we found the average effect for studies in hypocrisy conditions (k = 22 effect sizes from 22 sample clusters and 18 study clusters) represented a significant moral licensing effect (CEM Hedges' g = 0.25, 95% CI [0.09, 0.41]; HEM Hedges' g = 0.25, 95% CI [0.09, 0.42]) that is roughly equivalent in size to the 35th–40th percentile of social psychology effects (Lovakov & Agadullina, 2021).

We found that the average effect for studies in moral credits conditions (k = 132 effect sizes from 124 sample clusters and 97 study clusters) was not significantly different from zero (CEM Hedges' g = −0.03, 95% CI [−0.12, 0.05]; HEM Hedges' g = −0.04, 95% CI [−0.13, 0.04]). This null effect does not appear to be confounded by the area of literature, as even when only looking at non-FITD studies, moral licensing was not observed across all studies using positive manipulations (as we would expect based on our predictions; see Table 5).

One could argue that the valence of the target behavior is not a valid proxy for ambiguity, as some negative target behaviors (that we have operationalized as more blatant transgressions), may still be ambiguous. This is because some of these negative target behaviors are intentionally designed to be ambiguous by providing plausible alternative explanations for a transgression other than immorality. For instance, Monin and Miller (2001) conducted a study where participants had to determine if a job was more suitable for a White person over a Black person or for a man over a woman. This behavior was considered discriminatory and categorized as a blatant transgression in our analysis. The scenarios presented to participants were crafted in a way that made the employment setting appear hostile toward people of color or women. Consequently, the participants' choices could be attributed to prejudice or a desire to protect the Black or female character from a hostile work environment. Therefore, this target behavior aims to create high attributional ambiguity. If negative target behaviors of this nature are indeed better characterized as ambiguous, it could confound the results when using target behavior valence as a measure of ambiguity.

Despite our initial hesitance to code ambiguity according to criteria such as attributional ambiguity (whether the participants could plausibly perceive a legitimate justification for their behavior) due to the subjectivity in such a categorization system, we conducted post hoc coding to classify negative target behaviors according to whether they were high or low in attributional ambiguity. We

[3] For a breakdown of effect size estimates across all moderator subgroups included in Model 1, see additional online material (Table S6: https://osf.io/yb589).

Table 5
Mean Effect Sizes (Hedges' g) Within Combinations of Moderator Subgroups

Moderator subgroup	CEM estimate		HEM estimate	
	Prosocial-FITD	Non-FITD	Prosocial-FITD	Non-FITD
Negative target behaviors		0.25		0.24
Positive target behaviors	−0.43	0.03	−0.45	0.03

Note. Positive effect sizes represent a moral licensing effect; negative effect sizes represent a positive consistency effect. CEM = correlated effects model; HEM = hierarchical effects model; FITD = foot-in-the-door.

categorized studies using negative target behaviors in which a plausible alternative explanation for the behavior was possible as *high ambiguity* and those where this did not appear to be the case as *low ambiguity*. We present this post hoc coding in additional online material (Table S4: https://osf.io/yb589). Given the subjective nature of this coding approach, we also include the rationale for why each effect was categorized as high or low ambiguity in additional online material (Table S4: https://osf.io/yb589). All negative target behaviors were coded by two independent coders (Cohen's $\kappa = 0.81$). This resulted in 21 effect sizes using negative target behaviors being recoded as high ambiguity, 36 remained coded as low ambiguity, and 10 unclassified as insufficient information was available to code them.

We reclassified effect sizes into hypocrisy and moral credits conditions based on attributional ambiguity and domain-consistency. The average effect size for studies that met our conditions for hypocrisy ($k = 5$) was nonsignificant (CEM Hedges' $g = 0.01$, 95% CI [−0.43, 0.41]; HEM Hedges' $g = -0.01$, 95% CI [−0.61, 0.58]). Contrary to predictions of the moral credits model (but consistent with the analysis using target behavior valence), we also found a nonsignificant point estimate for studies that met our conditions for moral credits ($k = 148$; CEM Hedges' $g = 0.02$, 95% CI [−0.06, 0.09]; HEM Hedges' $g = 0.003$, 95% CI [−0.07, 0.09]).

Moral Credentials Conditions. To test the predictions of the moral credentials account that moral licensing should be at least observed when studies use ambiguous transgressions (operationalized as positive target behaviors) that are domain-consistent, we calculated the average effect size for studies that met these conditions ($k = 87$ effect sizes grouped into 80 sample studies and 63 study clusters). Contrary to our predictions, the average effect estimates for these studies represented significant positive consistency effects (CEM Hedges' $g = -0.16$, 95% CI [−0.26, −0.06]; HEM Hedges' $g = -0.16$, 95% CI [−0.26, −0.07]) equivalent in size to the 25th percentile for social psychology effects.

Like our test of the moral credits model, our test of the moral credentials model could be challenged on the basis that using target behavior valence does not accurately capture ambiguity, and our classification of effect sizes as meeting moral credentials conditions was based on this proxy. We therefore reclassified effect sizes according to whether they met conditions for moral credentials using our post hoc coding of attributional ambiguity and domain-consistency. Using this classification system, $k = 102$ studies met our conditions for moral credentials. The average effect size for these studies was nonsignificant (CEM Hedges' $g = -0.06$, 95% CI [−0.16, 0.03]; HEM Hedges' $g = -0.07$, 95% CI [−0.17, 0.03]).

Differences in Effects and Literature Over Time

We tested differences in effect sizes across time by examining whether publication year moderated effect sizes and whether there was a relationship between year and sample size. We then examined differences in when prosocial-FITD and non-FITD studies were published, and whether there were differences in sample size between them. We omitted Ebersole et al. (2016) from these analyses of sample size given the unusually large sample of this study could skew the results ($n = 3{,}286$). The moderation analyses indicated that effect sizes varied as a function of year (see Table 3). The results suggest that studies published in later years are more likely to produce moral licensing effects, and older studies are more likely to produce positive consistency effects. There was also a significant, positive relationship between publication year and sample size ($r = .33$, $p < .001$), which suggests sample sizes have become larger over time

Literature (prosocial-FITD and non-FITD) differed both in terms of publication date and sample size. All studies published prior to 2001 were prosocial-FITD studies (ranging from 1972 to 2018), while 79% of studies published from 2001 onward were non-FITD SMB studies (note that we select this date as a point of reference because it coincides with the date of the first published moral licensing study: Monin & Miller, 2001). Overall, the average sample size for prosocial-FITD studies ($M = 96.35$, $SD = 65.03$) was smaller and less variable than non-FITD studies ($M = 139.66$, $SD = 134.79$).

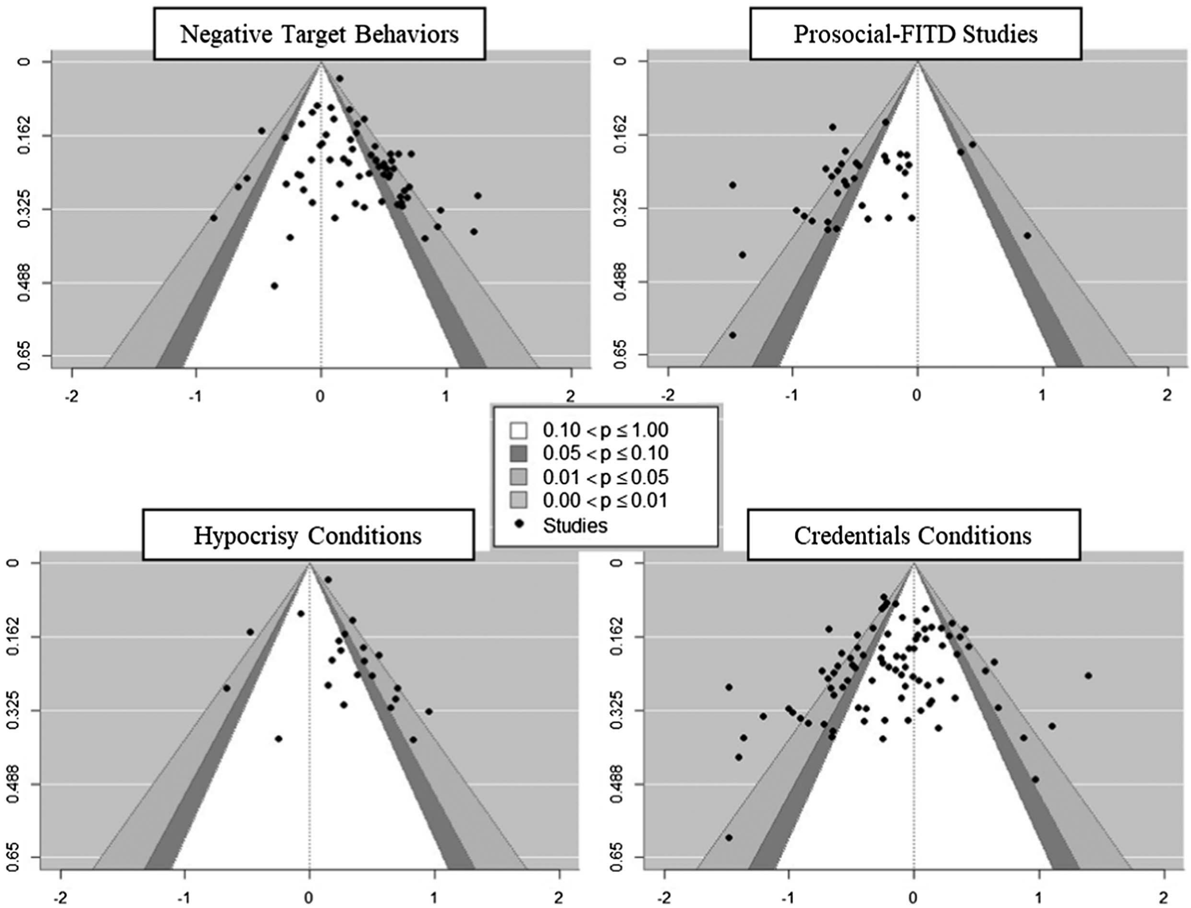
File-Drawer Bias

We evaluated the potential impact of file-drawer bias for conditions under which we found significant average effect sizes (e.g., negative target behaviors, prosocial-FITD studies, hypocrisy conditions, credentials conditions [note: we used our original categorizations based on target behavior valence]). We began by examining contour-enhanced funnel plots for studies within the aforementioned conditions (see Figure 3). Each plot is centered on the effect size point estimate for studies within those conditions, and the shaded areas indicate levels of significance. Across all plots, there appears to be a range of effect sizes distributed across both regions of significance and nonsignificance, which suggests an absence of file-drawer bias.

Next, we used PET-PEESE and 3-PSM analyses. The PET-PEESE analyses indicated there is a significant relationship between effect size and error for studies using negative target behavior, which is a possible indication of file-drawer bias (see β in Table 6). The analysis produced an adjusted effect size estimate that was

Figure 3

Contour Enhanced Funnel Plots (Meta-Analysis 1)



Note. Effect size (Hedges' *g*) is plotted on the *x*-axis; standard error is plotted on the *y*-axis. FITD = foot-in-the-door.

not significantly different from zero (Hedges' $g = -0.04$, 95% CI [−0.26, 0.18]). For the other subgroups, PET-PEESE did not suggest evidence of file-drawer bias, as the slope of the models suggested there was no systematic relationship between standard error and effect size. The 3-PSM model suggested that modeling file-drawer bias did not improve model fit for studies that met our operationalization of moral credentials conditions. Modeling bias did significantly improve model fit in all other conditions. Adjusted effect size estimates suggested that the point estimates for studies using negative target behaviors and studies that met our operationalization of hypocrisy conditions may not be significantly different from zero. For prosocial-FITD studies, modeling bias resulted in a larger effect size estimate, but the overlap in the confidence intervals for the original and adjusted estimates suggests these estimates may not be substantially different.

Discussion

In Meta-Analysis 1, we have presented the first quantitative synthesis of SMB studies using positive moral manipulations (k = 217 effect sizes, n = 31,242) and meta-analytic examination of methodological moderators of moral licensing and positive consistency. After controlling for overlap with other moderators, our analyses found evidence for both moral licensing and positive consistency, but under different methodological conditions. The results were consistent for both the CEM and HEM.

Conditions for Moral Licensing

Our results suggested that moral licensing only appears to reliably occur in studies that include negative target behaviors (see Tables 3 and 4). This was the case regardless of variation in other aspects of research design (e.g., domain-consistency, manipulation behavior type). That is, in these conditions, research participants who are first induced to imagine, recall, or perform a good deed appear to be more likely to engage in subsequent overtly negative behavior (e.g., cheating) but do not necessarily refrain from engaging in positive behavior. This was the case even if their subsequent negative behavior was in the same domain as their initial good behavior and so could be potentially perceived as directly contradicting it.

The average effect size for studies all using negative target behaviors (CEM Hedges' $g = 0.25$, 95% CI [0.16, −0.44]; HEM Hedges' $g = 0.24$, 95% CI [0.15, −0.33]) fell between the 35th–40th percentile of effects in social psychology (Lovakov & Agadullina,

Table 6
Publication Bias Analyses for Significant Moderator Categories

Moderator subgroup	PET-PEESE[a]	3-PSM	
	β [95% CI]	Fit improvement (χ^2)	SMg [95% CI]
Negative target behaviors	1.34 [0.35, 2.35]	25.05**	0.004 [−0.008, 0.09]
Prosocial-FITD studies	−0.22 [−1.80, 1.36]	4.58*	−0.47 [−0.59, −0.34]
Hypocrisy conditions	1.61 [−0.31, 3.53]	9.55**	0.01 [−0.15,0.17]
Credentials conditions	−0.74 [−1.72, 0.24]	2.17	−0.23 [−0.34, −0.12]

Note. Positive effect sizes indicate a moral licensing effect; negative effect sizes indicate a positive consistency effect. PET-PEESE = precision-effect test and precision-effect estimate with standard errors; 3-PSM = three-parameter selection model; CI = confidence interval; SMg = effect size corrected by 3-PSM; FITD = foot-in-the-door.
[a] PET estimates reported.
* $p < .05$. ** $p < .01$.

2021). It is possible that this result may overestimate the true size of the moral licensing effect in this condition, as both the PET-PEESE and 3-PSM model suggest that file-drawer bias may be present and produced adjusted effect size estimates that were not significantly different from zero. In line with recommendations from previous research (Carter et al., 2019; McShane et al., 2016), we recommend that the adjusted effect estimate provided by the 3-PSM models not be interpreted as estimates of the "true" effect sizes. Rather, we interpret the results as evidence that file-drawer bias is likely present and recommend that researchers assume the true effect sizes are smaller than those we report.

Conditions for Positive Consistency

The results of the metaregression (Table 4) indicated that positive consistency was only reliably observed in prosocial-FITD studies. The average effect for these studies (CEM Hedges' $g = -0.44$, 95% CI [−0.59, −0.26]; HEM Hedges' $g = -0.43$, 95% CI [−0.54, −0.33]) was roughly equivalent to the 55th–60th percentile of social psychology effects (Lovakov & Agadullina, 2021). This suggests that participants who are induced to agree to one request for help are subsequently more likely to agree to a second request for help. The PET-PEESE analysis (Table 6) and contour-enhanced funnel plots (Figure 3) did not suggest evidence of file-drawer bias, but the 3-PSM model suggested that file-drawer bias may be present. It is not unusual for different approaches to assessing file-drawer bias to produce different results (Carter et al., 2019), and so conclusions are typically made by evaluating multiple approaches. In this case, we interpret these results as evidence that we cannot reject the possibility that the effect size estimates produced by our analysis may overestimate the true effect size for prosocial-FITD studies.

Prosocial-FITD Versus Non-FITD Studies

As argued in the introduction, prosocial-FITD studies and SMB studies published outside this literature all use variations of the SMB design with positive manipulations and thus represent potential tests of moral licensing and positive consistency. Our analyses suggest that the area of literature predicted differences in effect sizes above and beyond other methodological differences between studies. That is, differences between areas of literature do not appear to simply be attributable to the fact that one (prosocial-FITD) tends to use positive target behaviors, while the other (non-FITD) uses a mix of positive and negative target behaviors, as we found that use of positive target behaviors was not sufficient to produce positive consistency in studies published outside of the prosocial-FITD literature. Indeed, when positive target behaviors were used in non-FITD studies, the estimated effect size was essentially zero.

These results suggest that there are differences in the specific methods used in prosocial-FITD and in non-FITD SMB studies not accounted for in these analyses that may explain why each tends to find opposite effects. Some notable differences we did were not able to test meta-analytically relate to (a) the type or domain of task involved in the different studies and (b) the difference in magnitude of task in the manipulation compared to the target behavior. Specifically, prosocial-FITD studies by design involve requests to *help* in some way, for example, by giving an experimenter direction (Saint-Bauzel & Fointiat, 2012) or helping promote a good cause by wearing a button (e.g., Chartrand et al., 1999). The domain of behavior may change, but the type of moral behavior (helping) remains the same. In contrast, the types of moral behaviors involved in non-FITD studies vary greatly regarding both the positive manipulations (e.g., imagine volunteering or engaging in eco-friendly behavior; Clot et al., 2018; Mazar & Zhong, 2010) and the target behaviors (e.g., donating to charity, cheating, self-indulgent food choices; Brown et al., 2011; Jordan et al., 2011; Khan & Dhar, 2006). In addition, the prosocial-FITD variation of the SMB design involves an *escalation* in the size of requests (a small initial request followed by a larger request), whereas this is not necessarily a pattern observed in non-SMB studies. We recommend future research investigate these differences to better understand the distinguishing features between prosocial-FITD and non-FITD SMB studies.

Each area of literature is likely to be differentially impacted by file-drawer bias. The prosocial-FITD literature is much older and so locating records of unreported studies in this field is much more difficult. Indeed, the unpublished studies included in our analyses were all recently conducted, non-FITD studies. The analysis of non-FITD studies may therefore be less impacted by file-drawer bias due to the inclusion of these unpublished studies (although we believe there are other unreported results in this field, we were unable to locate). This could account for the difference in the magnitude of the average effect sizes for prosocial-FITD studies and studies using negative target behaviors (all of which were non-FITD studies).

The reasons we have suggested for the variations between prosocial-FITD and non-FITD studies are speculative. Another possibility is that the differences stem from how researchers have labeled their studies over time. The FITD literature has a lengthy history, starting with the influential work by Freedman and Fraser (1966). It is plausible that moral licensing effects were indeed observed in those earlier years but remained unpublished due to being atypical compared to the rest of the literature. The publication of Monin and Miller's (2001) pioneering moral licensing study may have paved the way for the publication of results that contradicted previously published SMB studies.

Changes as a Function of Publication Year

Our analyses revealed some interesting changes in the literature over the years. First, publication year moderated effect sizes such that studies reported in later years were more likely to report moral licensing effects. This may reflect a shift in the *type of research* most frequently being conducted at different points in time, rather than just a tendency for research conducted in later years to just be more likely to find moral licensing. All studies published before 2001 were from the prosocial-FITD literature and so used a specific version of the SMB paradigm involving the presentation of sequential, escalating requests for engagement in prosocial behavior. After the publication of Monin and Miller's (2001) seminal research in 2001, there is a notable shift in the literature toward the use of other variations of the SMB paradigm (79% of studies published from this time onwards are non-FITD studies). As previously discussed, these studies use a much broader range of experimental tasks (including the use of negative target behaviors that are not used in prosocial-FITD studies). These differences in methodology may account for the difference in results between the literature, or this may also reflect a file-drawer bias for moral licensing studies before 2001.

It is not clear why prosocial-FITD studies have become less prevalent and other variations of SMB studies more prevalent after 2001. It seems reasonable to assume this may be simply because FITD effects had a long- and well-established history, having been studied since the 1960s (Freedman & Fraser, 1966), and researchers in later years were simply more interested in investigating novel research questions and paradigms. Alternatively, this could also reflect a file-drawer bias for moral licensing studies prior to 2001 as these results were anomalous compared to previous SMB studies. There is no evidence to suggest prosocial-FITD studies became more likely to find moral licensing, or less likely to find positive consistency, as we found publication year did not moderate effect sizes when just looking at prosocial-FITD studies ($Q_B = -0.02$, 95% CI [−0.03, 0.06]). This suggests that the effect of year that was initially observed across all included studies is best explained by the fact that there are more non-FITD SMB studies published in later years.

There was a positive relationship between sample size and year, which suggests sample sizes have increased over time. This may reflect a shift toward the widespread use of online research that is more conducive to recruiting larger samples (and less conducive to prosocial-FITD studies that typically involve inducing participants to engage in helping behavior). It could also suggest a greater emphasis on the importance of statistical power in later years. However, increased sample size and statistical power are unlikely to account for why later studies are more likely to find moral licensing, as the increase in sample size was observed for both prosocial-FITD and non-FITD studies.

Comparison to Previous Meta-Analyses

Previous meta-analyses have sampled only a portion of SMB studies using positive manipulations, and so it is unsurprising that the results of the current analyses are not entirely consistent with these past analyses. For example, both Blanken et al. (2015) and Simbrunner and Schlegelmilch (2017) found the average effect across all studies included in their analyses amounted to a significant moral licensing effect (Cohen's $d = 0.31$ and 0.32, respectively) roughly equivalent to the 45th percentile of social psychology effects (Lovakov & Agadullina, 2021). We found moral licensing effects of comparable in size only in studies that used negative target behaviors. This could suggest these types of SMB studies were overrepresented in previous meta-analyses relative to other SMB paradigm variations, due to the narrower selection criteria of these analyses (SMB studies that did not explicitly predict moral licensing were excluded). Our results are consistent with previous meta-analyses of FITD studies in that we found evidence for behavioral consistency in SMB studies that used the FITD paradigm (Beaman et al., 1983; Burger, 1999; Dillard et al., 1984; Fern et al., 1986; Pascual & Guéguen, 2005).

Theoretical Implications

The results of the current analyses do not fully align with the predictions of the theoretical models discussed earlier in this article. Below, we discuss how the results relate to each of the theories presented in the introduction.

Moral Credits. According to the moral credits model, moral licensing should only be observed in conditions in which participants are less likely to be concerned about appearing hypocritical (Effron & Monin, 2010; D. T. Miller & Effron, 2010). Specifically, researchers suggest that hypocrisy concerns are likely to occur when participants are presented with a target behavior that is (a) a blatant transgression and (b) in the same domain as their previous good deed. Outside of these conditions, moral licensing can occur. In the current meta-analysis, we tested this prediction using two approaches.

Our initial strategy involved categorizing hypocrisy and moral credits conditions based on the valence of the target behavior (which served as a proxy for ambiguity) and domain-consistency. We chose to operationalize ambiguity in this manner due to extensive prior research indicating an asymmetry in how people judge moral rules and transgressions related to acts of omission compared to acts of commission (Janoff-Bulman, 2012; Spranca et al., 1991; Yeung et al., 2022). Additionally, we believed that this approach provided a more objective coding method compared to other sources of ambiguity (e.g., attributional ambiguity), as ambiguity has only been directly manipulated in one SMB study (Brown et al., 2011). The results did not support the moral credits model's predictions. Instead, they revealed evidence of moral licensing in the hypocrisy conditions and no significant effect in the moral credits conditions.

Because there are studies intentionally utilizing negative target behaviors with high attributional ambiguity (Monin & Miller, 2001), there is a possibility that our initial analyses, using target behavior valence as a measure of ambiguity, may not accurately test the predictions of the moral credits model. Therefore, we performed a post hoc recoding of effect sizes from studies involving negative target behaviors, classifying them as either high or low in attributional ambiguity. This was not initially included in our analytic plan.

Regardless of the approach used, the results do not fully align with predictions of the moral credits model. While the nonsignificant effect we found for hypocrisy conditions is consistent with predictions of moral credits, the nonsignificant effect found for moral credits conditions was not. This suggests that the absence of hypocrisy concerns is not sufficient to prompt moral licensing, and other factors must account for when and why these effects can occur.

Moral Credentials. The moral credentials model predicts that moral licensing should only occur when the target behavior is (a) ambiguous and (b) in the same domain as the manipulation (Merritt et al., 2010; D. T. Miller & Effron, 2010). We operationalized moral credentials conditions first as studies that used positive target behaviors (as a proxy for ambiguity) that were in the same domain as the moral manipulation. Contrary to our predictions, we found evidence of positive consistency in these conditions. We then reclassified studies according to whether they met conditions for moral credentials or not based on domain-consistency and our post hoc recoding of attributional ambiguity. The results differed from our initial analysis, in that the average effect size for recoded moral credentials studies was nonsignificant; however, these results still do not align with predictions of the moral credentials model.

SPT. Based on the propositions of SPT that consistency should be more likely to occur if an initial behavior (a) involves more cost or effort to perform and (b) is in the same domain as target behavior (Bem, 1972; Burger, 1999), we predicted that manipulation behavior type and domain-consistency would moderate effect sizes. After controlling for overlap with other moderators, we found these methodological features did not moderate effect sizes. As such, predictions derived from SPT were unsupported.

Activation of Moral Identity. Effron and Conway (2015) argued that positive manipulations that focus on the behavior being performed should lead to moral licensing (i.e., behavior-focused manipulations), while those that make the moral identity of the person performing the behavior salient should lead to positive consistency (i.e., identity-focused manipulations). We used Effron and Conway's taxonomy as a guide when coding these moderator categories. The results indicated that the distinction between manipulations did not moderate effect sizes.

When coding this moderator category, we chose to classify any manipulation that induced participants to perform or recall a moral behavior without any explicit instructions to think abstractly or to focus on values, principles, or long-term goals, as behavior-focused. There is no way of knowing for certain whether participants were engaging in this kind of thinking independent of the instructions. As a stronger test of this prediction, we reanalyzed this moderator category using only studies that included specific instructions to think about behavior in a manner Effron and Conway (2015) argue will be less likely to activate moral identity (i.e., participants are explicitly instructed to think about consequences of the behavior or goal progress). The results remained unchanged even after limiting the category to these studies.[4] This suggests the broad taxonomy is not sufficient to explain when positive moral manipulations lead to moral licensing and sometimes to positive consistency.

Alternative Explanations. Van de Ven et al. (2018) have suggested an alternative explanation of moral licensing that may partially explain some of the results observed in our analyses. They argue that moral licensing may represent a form of motivated reasoning, in which people draw on their prior good deeds to justify indulging in a tempting, but morally problematic, behavior. That is, they characterize moral licensing as a conscious rationalization process.

In the present study, we found that moral licensing occurred only when studies measured engagement in negative target behaviors, which are arguably more likely to result in direct rewards (e.g., gains from cheating) than refraining from positive target behaviors, whose benefits represent conservation of resources or effort, rather than gains. That is, engaging in negative target behaviors may represent a greater temptation than refraining from positive target behaviors. This may create greater motivation to draw on past positive actions as an excuse to "do something bad."

This explanation is speculative, but it provides some insight into why negative target behaviors may be associated with moral licensing effects, whereas positive target behaviors were not, and it may provide a promising route forward for future research. For example, the claims of this account could be tested by measuring the relationship between the reward value of the target behavior and the likelihood of moral licensing occurring. This explanation does not account for why the prosocial-FITD variation of the SMB paradigm was associated with positive consistency effects. However, as argued in previous research, it is likely that consistency and licensing effects may be driven by different mechanisms (e.g., Mullen & Monin, 2016), and differences in methodological characteristics of prosocial-FITD studies may be responsible for the effect differences.

Lasarov and Hoffmann (2020) have recently proposed a model of social moral licensing that integrates the role of social processes into the moral credits and moral credentials models. For example, they argue that the likelihood of moral licensing is impacted by both (a) whether behavior is observed and (b) by whom it is observed. They argue that if behavior is observed by someone from the actor's in-group, licensing may occur, but not if it is observed by an out-group member. This is because the actor can assume an in-group observer is likely to interpret their behavior similarly to themselves (and thus excuse it), whereas this is not the case for an out-group observer. We did not test assumptions of this theory in the current analysis given it is not possible to determine whether participants perceive experimenters, confederates, or other participants observing their behavior as part of their in-group. Future research investigating the role of social factors in SMB effects may benefit from the knowledge provided by this meta-analysis regarding the methodological conditions under which moral licensing is most likely to be observed.

Practical Implications

The results have several practical implications for researchers studying SMB in general. The moderation analyses revealed that both moral licensing and positive consistency appear to occur, but under narrow conditions. Researchers wishing to further investigate mechanisms of moral licensing may benefit from the use of negative target behaviors (especially those high in attributional ambiguity), while researchers interested in positive consistency may benefit from the use of prosocial-FITD variation of the SMB paradigm.

[4] The average effect for behavior focused manipulations when limited to studies including specific instructions to focus on the consequences of behavior or progress toward goals (Hedges' $g = -0.04$, 95% CI [−0.26, 0.80], $k = 9$). This did not significantly differ from person-focused manipulations (Hedges' $g = 0.04$, 95% CI [−0.05, 0.13], $k = 55$), $Q_B(1) = 0.45$, $p = .621$.

Researchers should consider the magnitude of estimated effects when conducting power analyses. The effect size estimates for different methodological conditions provided in Table 5 may be useful to researchers for determining sample sizes under different conditions. We recommend that these values be considered upper estimates, as they may be inflated by file-drawer bias. Researchers may benefit from using more conservative effect size estimates in power analyses.

Although the PET-PEESE and 3-PSM models suggest the size of the moral licensing effect for studies using negative target behaviors may be accurate, we also acknowledge that the estimated effect size (CEM Hedges' $g = 0.25$, 95% CI [0.16, −0.44]; HEM Hedges' $g = 0.24$, 95% CI [0.15, −0.33]) is in contrast to the smaller estimate provided by Ebersole et al.'s (2016; Cohen's $d = 0.15$) large-scale replication of Monin and Miller (2001). Given the variability in type and domain of tasks used across the SMB literature, even among studies using negative target behaviors, we feel it prudent to recommend future research continue to conduct high-powered replications to provide precise effect size estimates of moral licensing in different contexts. To our knowledge, there are currently no high-powered replications of prosocial-FITD studies, and so we also encourage future research to conduct such replications. In the meantime, we recommend researchers assume small effects when conducting power analyses for research in this area.

Conclusions

The results of Meta-Analysis 1 indicate that significant moral licensing and positive consistency effects are observed, but in systematically different and narrow conditions. Moral licensing appears to occur reliably only in studies using negative target behaviors, while positive consistency appears to only occur in studies using sequential requests for prosocial behavior (i.e., prosocial-FITD studies). We recommend future research studying the effect of positive moral manipulations on subsequent moral behavior take these methodological conditions into account and consider the variable magnitude of effects in different conditions when conducting a priori power analyses for future research. We also recommend further research to investigate the mechanisms of positive SMB effects (both existing and new theories) to determine if those mechanisms can be integrated to explain the pattern of results observed in these analyses.

Meta-Analysis 2: Negative Moral Manipulations (Moral Compensation vs. Negative Consistency)

The SMB literature includes studies that examine the effect of engaging in a morally negative behavior on subsequent moral behavior (bottom half of Figure 1). These manipulations involve participants either engaging in, imagining, or recalling immoral actions such as intentionally harming another person (e.g., Greene & Low, 2014). Like studies using positive manipulations, negative SMB studies have the potential to find moral balancing (in this case referred to as *moral compensation*), consistency (in this case *negative consistency*), or null results. It is not clear under what conditions each effect is likely to occur. We review evidence for these effects below.

Evidence for Negative SMB Effects

Moral Compensation

Within the moral balancing literature, moral compensation effects have primarily been studied using tasks that ask participants to recall previous immoral or unethical actions (e.g., Cornelissen et al., 2013; Jordan et al., 2011). These studies have found that such recall tasks can lead to a greater likelihood of volunteering to help others (Young et al., 2012), higher prosocial intentions (Jordan et al., 2011), less cheating (Cornelissen et al., 2013; Jordan et al., 2011), and making fewer immoral decisions in hypothetical scenarios (Greene & Low, 2014). Similar effects have been found in the literature exploring guilt and compliance, a subset of which uses experimental manipulations in which participants are induced to harm (Carlsmith & Gross, 1969) or to think they have harmed (Konecni, 1972) another person before being given the opportunity to engage in subsequent moral behavior. For example, Carlsmith and Gross (1969) found that participants who were led to believe they were administering painful electric shocks to another person were subsequently more likely to volunteer to help an environmental campaign than those participants who did not believe they had harmed anyone.

Compensatory effects have also been observed in the DITF literature. Like the FITD strategy, DITF is a compliance technique that involves the presentation of sequential requests. The first request is large enough to elicit refusal (i.e., refusal to help another person; Cialdini et al., 1975) and is followed by a second, smaller request. As was the case for FITD studies, DITF studies using prosocial requests (e.g., requests to volunteer for blood drives; Henderson & Burgoon, 2013) use a variation of the SMB paradigm. We refer to these as *prosocial-DITF* studies. In this case, refusal of the first request can be interpreted as an act of omission (failure to help), while the second request represents a second opportunity to help (a positive target behavior). Meta-analyses of the DITF paradigm have found that, compared to control, participants who refuse an initial request are more likely to agree to a second request (Feeley et al., 2012; O'Keefe & Hale, 1998). These meta-analyses also show that the DITF effect is typically larger for prosocial requests than other requests.

Like positive SMB studies, there have also been examples of negative SMB studies that have failed to find significant effects. Several studies using recall manipulations have found no difference in subsequent moral behavior between negative manipulation and control conditions (Hayley & Zinkiewicz, 2013; Young et al., 2012). In addition, there have been studies in which participants have either been induced to commit an immoral action (e.g., to cheat; Ploner & Regner, 2013) or to think that they have done so (e.g., participants told they have had a greater negative impact on the environment than others; Ho et al., 2015) that have found no effects on subsequent moral behavior (generosity or environmentally friendly behavior, respectively). Finally, a high-powered replication of Sachdeva et al. (2009) and Blanken et al. (2014) found no difference in charitable donations or decisions in a hypothetical moral dilemma between participants who had previously written about having immoral traits and participants in the control condition.

Negative Consistency

Negative consistency refers to the increased likelihood of engaging in a negative target behavior (either more immoral or less moral behavior) after an initial negative moral manipulation. These effects are much less prevalent in the literature, but there are a few notable examples (e.g., Conway & Peetz, 2012; Cornelissen et al., 2013; Gino et al., 2010; Henderson & Burgoon, 2013; Lee et al., 2016). For example, Cornelissen et al. (2013) found that participants asked to focus on abstract moral values when recalling a past unethical behavior subsequently gave less in a dictator game and were more likely to cheat than participants who recalled a neutral behavior. In contrast, they found the opposite results for participants who thought about their past unethical actions concretely (i.e., they thought about the immediate consequences of the behavior). Within the DITF literature, Henderson and Burgoon (2013) have also found that participants primed to think abstractly in a prosocial-DITF experiment were more likely to refuse sequential requests to assist with a blood donation campaign than participants primed to think concretely.

Theoretical Accounts of Negative SMB Effects and Theoretical Moderators

Theoretical explanations for moral compensation and negative consistency are similar to the explanations of moral licensing and positive consistency. Moral compensation has been explained via moral credits (Jordan et al., 2011; Sachdeva et al., 2009; Zhong et al., 2010), while negative consistency is explained via SPT (Cornelissen et al., 2013; Henderson & Burgoon, 2013). As such, these theories may provide a basis for testing moderators of negative SMB effects in the current analysis.

Moral Credits: An Account of Moral Compensation

The moral credits model proposes that moral compensation occurs when engaging in an initial morally negative action causes a person's moral self-image to fall below the equilibrium point (Nisan, 1990, 1991; Zhong et al., 2010). Subsequent positive behaviors provide a mechanism to repair moral self-image by returning moral self-image to a comfortable level.

Moderators Based on Moral Credits. Unlike the case of moral licensing, the moral credits model does not predict conditions under which moral compensation will not occur. Rather, it is assumed that any experimental manipulation that negatively affects moral self-image should prompt compensatory action to repair self-perceptions. Based on this model, we can predict that the degree of the initial transgression should affect the extent of observed compensatory action. This is because a more severe transgression should cause greater damage to moral self-image, and so more compensatory behavior is needed to repair that damage (Nisan, 1990, 1991). We tested this prediction by examining whether effect sizes differed as a function of type of transgression involved in the moral manipulation. Evidence for the omission–commission asymmetry suggests people judge acts of commission (i.e., engagement in negative behavior, such as harming someone) as more immoral and intentional than acts of omission (i.e., failure to engage in positive behavior, such as refusing to help someone; e.g., Bostyn & Roets, 2016; Kordes-de Vaal, 1996; Spranca et al., 1991; Yeung et al., 2022). This led us to predict that effect sizes in this meta-analysis would be larger for the former type of manipulations, even though moral compensation effects are expected to occur under both conditions.

SPT: An Account of Negative Consistency

Like positive consistency, negative consistency effects have largely been attributed to self-perception processes (Conway & Peetz, 2012; Cornelissen et al., 2013; Henderson & Burgoon, 2013; Lee et al., 2016). SPT proposes that people infer information about their character, identity, or behavioral preferences from their past actions and then act in a manner that is consistent with these self-perceptions when later presented with a similar choice (Bem, 1972). In the case of negative consistency, it has been argued that past behavior may signal to participants that adhering to moral values is not particularly important to them (Cornelissen et al., 2013), or they prefer self-interested, rather than prosocial behavior (Henderson & Burgoon, 2013; Lee et al., 2016). Thus, subsequent behavior may be negative as it is influenced by these negative self-signals.

Moderators Based on SPT. The same conditions that constrain positive consistency effects should also constrain negative consistency effects. That is, negative consistency should only occur when the manipulation and target behavior are in the same domain. Consistency should also be more likely to occur when the manipulation is more costly and effortful (i.e., involves real behavior) than when it is less so (i.e., recall or hypothetical manipulations). As such, in the current analysis, domain-consistency and type of behavior used in the moral manipulation were examined as moderators.

Exploratory Moderators. As in Meta-Analysis 1, the type of target behavior (real vs. hypothetical) and publication year were also tested as exploratory moderators. Unlike Meta-Analysis 1, the area of literature (moral compensation, guilt and compliance, and DITF) was not tested as a moderator because these areas of literature do not make systematically different predictions about the direction results. We did not measure whether inducing participants to believe the manipulation and target behavior would be observed by one or two audiences would moderate results as we did in Meta-Analysis 1, as this was examined as a potential source of hypocrisy concerns that are not relevant to negative SMB effects.

Method

The search and screening for Meta-Analysis 2 was conducted concurrently with that of Meta-Analysis 1, using the same search terms (see additional online material, Table S1: https://osf.io/yb589 for terms). Meta-Analysis 2 used the same (a) study inclusion criteria, (b) data extraction methods, and (c) meta-analytic procedures as Meta-Analysis 1, except for some minor differences explained below.

Inclusion Criteria

The inclusion criteria were identical to that used in Meta-Analysis 1; however, in this case, the moral manipulation had to induce people to recall, imagine, or perform a morally *negative* behavior, rather than a positive one.

Data Extraction and Moderator Coding

Data extraction procedures were the same as Meta-Analysis 1. Moderator coding differed only in that we did not code the area of literature or the number of audiences for negative SMB studies, and we included the additional moderator category of manipulation transgression type (omission or commission). Some manipulations could not be categorized in this way (i.e., recall tasks that just ask participants to recall acting immorally, and so the actual behavior they recall is unknown) and so were not included in the analysis of this moderator.

Meta-Analytic Procedure

Effect Size Calculations. As in Meta-Analysis 1, Hedges' *g* was selected as the effect size metric for the analysis. Effect sizes were coded so that positive values represent a moral balancing (in this case *moral compensation*, i.e., increased positive behavior after a negative manipulation), while negative values represent moral consistency (in this case *negative consistency* effect, i.e., decreased positive behavior after a negative manipulation).

Point Estimate Calculations. Point estimates (i.e., average effect sizes) were calculated by combining individual effect sizes used a random-effects model using RVE. As in Meta-Analysis 1, we calculated results using both a CEM and a HEM to account for dependency in the data.

Outliers. We ran influence diagnostics using the *metafor* package in *R* to identify potential outliers. Cases with studentized residuals >±1.96 were considered outliers, and cases with DFFITs values >±0.26 (i.e., $3 \times \sqrt{\frac{p}{k-p}}$) were considered influential (Viechtbauer & Cheung, 2010).

Moderation Analyses. Moderators were dummy-coded (see Table 7) and tested using metaregression. This was conducted using both CEM and HEM weights. Interrater agreement (Cohen's κ) for moderator coding is presented in Table 8.

File-Drawer Bias. As in Meta-Analysis 1, we assessed file-drawer bias using visual inspection of contour-enhanced funnel plots. We also used PET-PEESE and 3-PSM to test for publication bias and calculate adjusted effect size estimates.

Results

Study Characteristics

We extracted $k = 132$ effect sizes for Meta-Analysis 2. Seven effect sizes were identified as outliers based on their studentized residual values. The DFFITs values for these cases were not problematic and so these cases were retained for the analyses. The 132 effect sizes were grouped into 98 study clusters and 122 sample clusters. The total sample size was $N = 14{,}443$.

Descriptive statistics for Meta-Analysis 2 are presented in Table 8. The most frequently used target behaviors were positive (n = 113, vs. n = 18 negative), in the same domain as the manipulation ($n = 58$, vs. $n = 48$ domain-inconsistent), and measured hypothetical behavior (n = 73, vs. n = 59 actual behavior). Most manipulations were behavioral (n = 84), followed by recall (n = 29), and hypothetical (n = 19). Behavior-focused manipulation (n = 91) was more frequently used than identity-focused manipulation (n = 36), and manipulations were frequently involved acts of commission (n = 64) compared to acts of omission (n = 43). Nearly all effect sizes were extracted from published journal articles (n = 127), with only a small number extracted from unpublished sources (n = 5).

Sample sizes ranged considerably between studies (18–1,067). Average sample sizes within some moderator categories were somewhat skewed due to the inclusion of one particularly large sample (n = 1,067; Guéguen, 2003), but the median sample size was similar across most moderator subgroups. There was some variation in the average publication year between some subgroups. For example, there was a trend toward earlier publication year for studies using positive target behaviors (M = 1991.69), that were domain-inconsistent (M = 1991.98) or used behavioral manipulations (M = 1994.29) compared to those using negative target behaviors (M = 2012.83), that were domain-consistent (M = 2003.07), and that use hypothetical or recall manipulations (*M*s = 2011.58, 2015.00, respectively).

Search and Screening

The results of the search and screening process for Meta-Analysis 2 are reported in Figure 2. A summary of individual effect sizes, study descriptions, and moderator coding can be found in the additional online material (Table S5: https://osf.io/yb589).

Point Estimate Across All Effect Sizes

Using both CEM and HEM estimates, we found the average effect across negative SMB studies was a significant moral compensation effect CEM Hedges' g = 0.23, 95% CI [0.15, 0.32], p < .001, τ = 0.3; HEM Hedges' $g = 0.22$, 95% CI [0.14, 0.31], $p < .001$, $\tau = 0.36$ (approximately 35th percentile for effects in social psychology; Lovakov & Agadullina, 2021).

Moderator Analyses

Moderators were tested using metaregression. Table 7 presents effect size estimates for studies within each moderator subgroup (Hedges' g) and metaregression coefficients (Q_B). Significant regression coefficients indicate that effect sizes differ systematically as a function of that moderator. The results of the CEM model revealed effect sizes were moderated by the target behavior's valence, but the HEM model suggested this might not be the case. For both the CEM and HEM models, the average effect size for studies using positive target behaviors was a significant moral compensation effect (CEM Hedges' g = 0.27, 95% CI [0.18, 0.37]; HEM Hedges' $g = 0.26$, 95% CI [0.17, 0.35]), which was equivalent in size to the 40th percentile for social psychology effects (Lovakov & Agadullina, 2021). In contrast, the average effect size for studies using negative target behaviors was not significantly different from zero. No differences in effect sizes were found for any other moderators that we tested.

File-Drawer Bias

We evaluated file-drawer bias both across all included studies and just studies that used positive target behaviors. Contour-enhanced funnel plots are presented in Figure 4. Both plots appear extremely similar, as a substantial proportion of included studies (~86%) used positive target behaviors. In both bases, the plots showed a relatively symmetrical pattern with points dispersed across zones of significance. This suggests the absence of file-drawer bias.

Conditional PET-PEESE (PET estimates reported) indicated that there was a systematic relationship between effect sizes and standard errors when looking at all studies (β = 1.21, 95% CI [0.41, 2.00]; adjusted Hedges' g = −0.08, 95% CI [−0.29, 0.14]) and also when looking just at studies using positive target behaviors ($\beta = 1.10$, 95% CI [0.32, 1.88]; adjusted Hedges' $g = -0.01$, 95% CI [−0.23, 0.21]).

Table 7

Meta-Analysis 2: Average Effect Size and Heterogeneity Estimates for Full Sample and Moderator Categories

		Correlated effects RVE model					Hierarchical weights RVE model				
Moderator	k	SmC	g	95% CI	τ	Q_B [95% CI]	StC	g	95% CI	τ	Q_B [95% CI]
Target behavior valence						**0.30 [0.03, 0.57]**					0.28 [−0.003, 0.57]
Negative (0)	18	17	−0.02	[−0.28, 0.24]	0.44		15	−0.02	[−0.29, 0.26]	0.37	
Positive (1)	113	104	**0.27**	**[0.18, 0.37]**	0.36		83	**0.26**	**[0.17, 0.35]**	0.35	
Domain-consistency						−0.07 [−0.29, 0.15]					−0.05 [−0.27, 0.18]
Inconsistent (0)	48	46	**0.30**	**[0.11, 0.49]**	0.49		35	**0.28**	**[0.10, 0.47]**	0.42	
Consistent (1)	58	53	**0.22**	**[0.10, 0.34]**	0.38		44	**0.22**	**[0.10, 0.36]**	0.20	
Manipulation behavior type											
Contrast 1: Hypothetical (0) versus recall (1)						−0.003 [−0.24, 0.23]					−0.04 [−0.26, 0.18]
Contrast 2: Hypothetical (0) versus behavior (1)						0.05 [−0.17, 0.28]					0.05 [−0.18, 0.29]
Hypothetical	19	19	**0.19**	**[0.002, 0.38]**	0.27		16	0.19	[−0.001, 0.38]	0.23	
Recall	29	24	**0.20**	**[0.05, 0.34]**	0.28		20	**0.16**	**[0.04, 0.28]**	0.38	
Behavioral	84	79	**0.26**	**[0.14, 0.39]**	0.44		62	**0.26**	**[0.13, 0.39]**	0.41	
Manipulation focus						0.05 [−0.15, 0.25]					0.04 [−0.16, 0.23]
Identity-focus (0)	36	33	**0.18**	**[0.02, 0.34]**	0.30		32	**0.19**	**[0.04, 0.34]**	0.41	
Behavior-focus (1)	91	89	**0.24**	**[0.13, 0.34]**	0.40		69	**0.23**	**[0.12, 0.34]**	0.40	
Target behavior type						−0.14 [−0.31, 0.04]					−0.13 [−0.30, 0.04]
Hypothetical (0)	73	66	**0.30**	**[0.18, 0.42]**	0.43		52	**0.29**	**[0.16, 0.41]**	0.33	
Behavior (1)	59	56	**0.16**	**[0.03, 0.29]**	0.33		45	**0.15**	**[0.03, 0.27]**	0.32	
Type of transgression						−0.02 [−0.23, 0.19]					−0.04 [−0.26, 0.19]
Omission (0)	43	38	**0.27**	**[0.11, 0.44]**	0.42		31	**0.27**	**[0.09, 0.45]**	0.27	
Commission (1)	64	61	**0.25**	**[0.12, 0.39]**	0.39		46	**0.23**	**[0.10, 0.37]**	0.38	
Year						**−0.01 [−0.01, −0.003]**					**−0.01 [−0.01, −0.003]**

Note. Significant effect sizes and regression coefficients are in boldface. RVE = robust variance estimation; k = total effect sizes; SmC = sample clusters; g = Hedges' g average effect size estimates for studies within moderator subgroup (positive values represent a moral compensation effect; negative values represent a negative consistency effect); CI = confidence interval; Q_B = regression coefficients for tests of moderators; StC = study clusters.

Table 8
Descriptive Statistics for Negative Sequential Moral Behavior Studies

Moderator	*k*	Sample size *M*	*SD*	*Mdn*	Range	Publication year *M*	*SD*	*Mdn*	Range
All studies	132	109.72	129.67	72	18–1,067	2001.33	17.56	2011	1966–2022
Target behavior valence (κ = 1.00)									
Negative	18	76.78	42.37	71.5	34–220	2012.83	2.33	2012.5	2010–2018
Positive	113	89.25	138.45	73	18–1,067	1999.69	18.21	2009	1966–2022
Domain-consistency (κ = .82)									
Inconsistent	48	70.25	28.22	56	22–355	1991.98	18.95	1985.5	1966–2019
Consistent	58	129.21	157.10	84.5	22–1,067	2003.07	15.73	2011.5	1967–2021
Manipulation behavior type (κ = .89)									
Hypothetical	19	114.16	131.57	77	18–634	2011.58	2.57	2012	2008–2016
Recall	29	126.38	107.22	90	29–356	2015.00	3.32	2014	2010–2022
Behavioral	84	102.49	136.93	64	22–1,067	1994.29	18.49	2000.5	1966–2021
Manipulation focus (κ = 1.00)									
Identity-focused	36	134.94	192.92	77.5	18–1,067	2000.56	14.91	2004.5	1975–2021
Behavior-focused	91	94.29	90.13	66	22–558	2000.92	18.72	2012	1966–2020
Target behavior type (κ = .92)									
Hypothetical	73	105.30	96.42	72	22–558	1999.53	17.41	2009	1967–2019
Behavior	59	114.51	162.44	72	18–1,067	2003.54	17.63	2012	1966–2022
Type of transgression (κ = .96)									
Omission	43	140.56	179.71	82	40–1,067	2000.04	15.02	2003	1975–2021
Commission	64	85.06	93.05	64	22–634	1997.59	19.91	2011	1966–2022
Publication status (κ = 1.00)									
Unpublished	5	122.4	45.70	98	90–196	2015.20	3.56	2014	2012–2019
Published	127	108.91	131.94	70	18–1,067	2000.78	17.67	2010	1966–2022

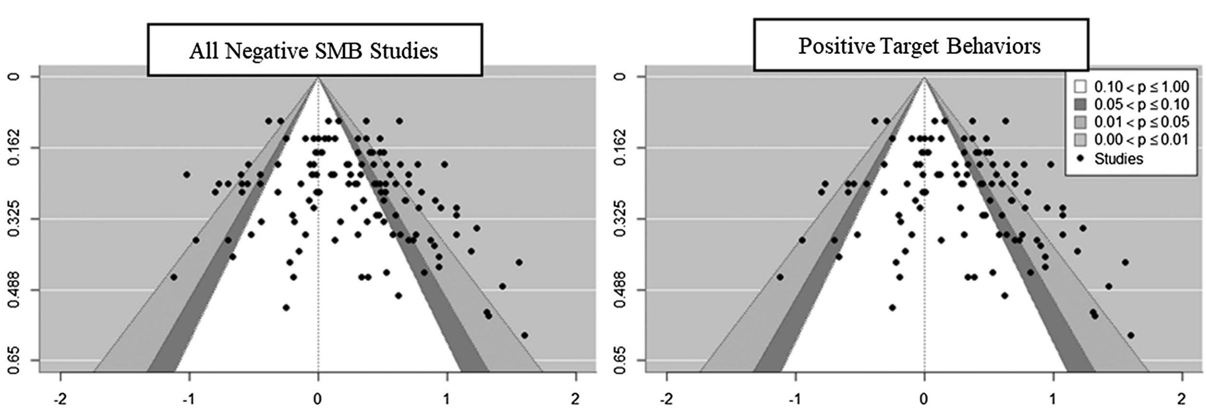
Modeling publication bias with 3-PSM significantly improved model fit when analyzing all included studies ($\chi^2 = 6.05$, $p = .014$, adjusted Hedges' $g = 0.10$, 95% CI [−0.02, 0.22]) and for studies using positive target behaviors ($\chi^2 = 7.39$, $p = .007$; adjusted Hedges' $g = 0.12$, 95% CI [−0.0001, 0.24]). The revised effect size estimates for both PET-PEESE and 3-PSM suggest the effects may be nonsignificant.

Discussion

The aim of Meta-Analysis 2 was to synthesize the results of SMB studies using negative moral manipulations and evaluate conditions under which moral compensation and negative consistency may occur. The average effect of these studies ($k = 132$ effect sizes, $N = 14{,}443$) was a significant moral compensation effect (CEM Hedges' $g = 0.23$, 95% CI [0.15, 0.32], HEM Hedges' $g = 0.22$, 95% CI [0.14, 0.31]) roughly equivalent to the 35th percentile of effects observed in the social psychology (Lovakov & Agadullina, 2021). This suggests that, on average, negative manipulations result in compensatory action (i.e., an increase in subsequent positive behavior), rather than prompting further negative behavior. Moderation analyses suggest this may only occur when the subsequent target behavior is positive and not when it involves a negative behavior (see Table 7).

PET-PEESE and 3-PSM analyses suggest that both the point estimates for all included studies and the point estimate for studies using positive target behaviors may be inflated by file-drawer

Figure 4
Contour Enhanced Funnel Plots (Meta-Analysis 2)



Note. Effect size (Hedges' g) is plotted on the x-axis; standard error is plotted on the y-axis. SMB = sequential moral behavior.

bias. In fact, both analyses suggest that these point estimates may be zero. Previous research suggests that both PET-PEESE and 3-PSM can overestimate the impact of file-drawer bias when heterogeneity is high and study sample sizes are small (e.g., Stanley, 2017), both of which were the case here. We interpret the results of our analyses of file-drawer bias as evidence that our meta-analytic estimates are inflated by file-drawer bias, but we do not recommend interpreting corrected effect sizes as estimates of the true moral compensation effect.

Theoretical Implications

The results of Meta-Analysis 2 do not fully align with either of the theoretical explanations for negative SMB that we discussed. We discuss these points below.

Moral Credits. Based on the moral credits account, we predicted that moral compensation should occur whenever participants engage in a negative manipulation, as this will cause a decrease in moral self-image (Nisan, 1990, 1991; Zhong et al., 2010). The difference in effect sizes observed between studies using positive and negative target behaviors is not explained by this account. It could be argued that refraining from negative behavior is not sufficient to repair moral self-image, and so only subsequent moral behavior provides an opportunity for compensation. However, this is a post hoc interpretation and not derived from theoretical explanations for moral compensation.

The moral credits accounts also predict that the size of the initial moral transgression (and by extension, the degree of impact on moral self-image or degree of guilt) should differentially affect the degree of compensatory behavior observed (Jordan et al., 2011). We therefore reasoned that effect sizes should differ between studies using negative manipulations that involve acts of commission compared to acts of omission (as the former is typically judged as a greater transgression than the latter; Janoff-Bulman et al., 2009; Spranca et al., 1991). Contrary to expectations, the type of transgression in the manipulation did not moderate effect sizes.

SPT. Based on SPT, we predicted that negative consistency should be most likely to occur (a) when the manipulation and target behavior are domain-consistent and (b) when the manipulation involves the performance of actual behavior. This is because past behavior can only be used as a heuristic for future decisions if the two behaviors are conceptually related, and behavioral manipulations are more effortful and thus more diagnostic of the self than hypothetical manipulation (Bem, 1972). The results of the current analysis did not support an SPT account, as we found evidence of reliable negative consistency effects in any conditions.

Alternative Explanations. The moral credits model and SPT are not the only explanations for negative SMB effects. Prosocial-DITF effects have been explained using a range of mechanisms such as reciprocal concessions (Cialdini et al., 1975), perceptual contrast (R. L. Miller et al., 1976), and self-presentation concerns (Pendleton & Batson, 1979). These mechanisms are not easily applied to experimental tasks that do not involve sequential requests for help and so were not included in this analysis.

Another explanation that has been used both in the prosocial-DITF literature and the guilt and compliance literature is that compensatory effects are driven by guilt (e.g., Boster et al., 2016; Freedman et al., 1967; O'Keefe & Figgé, 1997). This theoretical perspective was not explored in this analysis, as it is not possible to distinguish this explanation from moral credits without direct measurement of the mediating role of guilt (which was not possible in this analysis). As this analysis has not provided equivocal support for the models that were tested, future research would benefit from testing the utility of other theoretical explanations for explaining moral compensation effects.

Practical Implications

The results of Meta-Analysis 2 suggest that moral compensation does not reliably occur in studies using negative target behaviors. Researchers investigating the mechanisms of moral compensation should consider these findings when selecting appropriate target behaviors for their research. Researchers should also consider the magnitude of observed effects (Hedges' $g = 0.27$) when conducting power analyses. Given the potential presence of file-drawer bias, we encourage researchers to consider this an upper estimate and consider more conservative effect size estimates when determining appropriate sample size and power. As with our recommendations in Meta-Analysis 1, we again encourage researchers to conduct high-powered replications of these effects to obtain more precise estimates of effects for studies using specific experimental tasks.

Conclusions

The results of Meta-Analysis 2 suggest that the average effect size of SMB studies using negative manipulations was a significant moral compensation effect. This appears to be driven by the results of studies using positive target behaviors. In contrast, the average effect size for studies using negative target behaviors was null. The size of the average effect for all studies, and the effect for studies using positive target behaviors, may be inflated by file-drawer bias.

General Discussion

In this article, we have presented the first meta-analyses synthesizing the results of SMB studies examining the effect of positive (Meta-Analysis 1) and negative (Meta-Analysis 2) moral manipulations on subsequent moral behavior. These analyses have provided valuable insights into the methodological conditions under which specific SMB effects are most likely to be observed.

Meta-Analysis 1 revealed that moral licensing was only observed when participants were given the opportunity to engage in negative target behaviors (e.g., cheating). Using Cohen's (1988) rules of thumb for interpreting effect sizes, this would be considered a small effect (Hedges' $g = 0.25$). Relative to other effects reported in social psychology, this may be more appropriately interpreted as a moderate effect (35th–40th percentile for effects in this field; Lovakov & Agadullina, 2021). We also found evidence for positive consistency effects, but only for prosocial-FITD studies, and not for other variations of SMB studies using positive manipulations. The magnitude of this effect (Hedges' $g = -0.43$) is approximately moderate according to Cohen's (1988) guidelines but above-moderate (55th–60th percentile) according to Lovakov and Agadullina's (2021). Meta-Analysis 2 revealed that moral compensation occurred only when participants were given the opportunity to engage in a positive target behavior (e.g., donating to charity). The size of this effect (Hedges' $g = 0.29$) was comparable in magnitude to the moral licensing effect observed for negative target behaviors.

To provide additional context for interpretation, the moral balancing effects we observed are significantly smaller than effect sizes found in other meta-analyses that explore relationships between stable traits and moral behavior, but the effect size we discovered for positive consistency is roughly similar. For instance, previous meta-analyses have identified correlations between moral reasoning and moral behavior, and moral reasoning and immoral behavior, that are approximately equivalent to Hedges' $g = 0.48$ and 0.38, respectively (Wu & Liu, 2014). Additionally, correlations between moral identity and moral behavior were found to be equivalent to Hedges' $g = 0.45$ (Hertz & Krettenauer, 2016), and correlations between moral development and altruism were equivalent to Hedges' $g = 0.43$ (Villegas de Posada & Vargas-Trujillo, 2015).

We suggest the effect size estimates we have calculated be interpreted as upper estimates given the possibility that they are inflated by file-drawer bias. We encourage future researchers to consider both the methodological conditions under which each effect was observed and the estimated size of effects under these conditions when designing future research. We also emphasize the need for high-powered replications to obtain precise estimates of effects using specific experimental tasks.

Limitations of the Literature

Meta-analyses are only as good as the studies included in them. We controlled for some aspects of study quality by only including studies that used random allocation to condition and included neutral control groups. We acknowledge several limitations in the quality of studies included in the analyses. First, based on our effect size estimates, it is clear that many of the studies included in these analyses are underpowered. Low power can lead to overestimation of effect sizes in empirical studies, and significant results found in low-powered studies often do not replicate (Button et al., 2013). This may help explain why several prominent studies in the field have recently failed to replicate (e.g., Mazar & Zhong, 2010; Sachdeva et al., 2009) or why the successful replication of Monin and Miller (2001) resulted in a smaller effect size estimate than the original study (Ebersole et al., 2016).

There was substantial variation in the type and domain of tasks included in studies (both in terms of manipulation tasks and target behaviors). This is problematic given that morality is inherently subjective, and it is possible some tasks may not have been perceived as morally salient by all participants. For example, some studies operationalized moral behavior as pro-environmental behavior (e.g., Geng et al., 2016; Mazar & Zhong, 2010), yet environmental behavior may not be moralized by all people (e.g., Salomon et al., 2017). The moral credits model assumes that moral licensing and moral compensation occur because morally salient behavior impacts moral self-image—but if behavior is not perceived as morally salient, this would not be expected. Studies in this area tend to assume behavior is perceived in moral terms, rather than empirically verifying this by piloting experimental tasks or including manipulation checks. We also acknowledge that this means readers may dispute the inclusion of some tasks included in the analyses based on whether they truly constitute moral behaviors. The variability in experimental tasks also likely contributed to the high heterogeneity observed in the analyses, this suggests that even in conditions under which significant effects were observed, there is notable variation in the studies used to calculate those effects.

Limitations of the Current Analyses

There are also limitations to the approach taken for the current analyses. First, we feel it is highly likely that there are SMB studies relevant to this review that were not located by our search strategy. This is because the literature uses inconsistent terminology to describe experimental procedures and types of effects in this field. Having said that, given the breadth of our search and substantial number of studies screened and included in the analysis, we are confident that the literature included in this analysis, though it might not be exhaustive, provides a broad, representative sample of studies in this field.

In terms of moderator coding, we acknowledge that the coding of some moderators required subjective judgments as they pertained to factors not directly addressed in study reports. For example, coding factors such as domain-consistency and manipulation focus required interpretation of methodology by the authors. A random subset of the data was double-coded, and interrater agreement was high. However, due to practical constraints, it was not possible to double code moderator classification for all studies. We also acknowledge that our operationalization of target behavior ambiguity in terms of the target behavior's valence is only one possible interpretation of it. The post hoc recoding of negative target behaviors based on attributional ambiguity may partially address concerns with our original coding strategy, but the limited number of studies categorized as meeting hypocrisy conditions using this approach means it may not sufficiently test the predictions of the moral credits model. Consequently, the overall support for this model in the literature remains uncertain, and further empirical research is needed to directly address these questions.

There is currently no gold-standard approach to assessing file-drawer bias, and so we followed recommendations to use multiple approaches (Carter et al., 2019; McShane et al., 2016). The two correction methods used, PET-PEESE and 3-PSM, both have shortcomings and lead to inconsistent results at times. Previous research has shown that both PET-PEESE and 3-PSM can overestimate file-drawer bias and in turn overcorrect effect sizes when heterogeneity is high and sample sizes of individual studies are small (Carter et al., 2019; Stanley, 2017). Although it may be overly pessimistic to dismiss effects corrected to nonsignificant by the 3-PSM model without further investigation, we cannot reject the possibility that our estimates are inflated by file-drawer bias. To that end, we reiterate our call for high-power replication projects to investigate the replicability of specific effects and estimate effect sizes in the specific methodological conditions identified as moderating the results.

Most of the literature in this field has been conducted in Western countries, using samples of primarily Western participants. The overrepresentation of Western samples in research limits the generalizability of findings beyond the broad Western cultural context (Henrich et al., 2010). Although there is little SMB research from non-Western samples, Simbrunner and Schlegelmilch's (2017) meta-analysis of moral licensing found some cross-cultural variation in effects (moral licensing was more likely in Western samples, and positive consistency was more likely in South-Eastern samples). However, only $k = 5$ non-Western samples were included in their analyses. Given the limited number of non-Western studies available in the SMB field, we did not analyze cross-cultural differences in the current article. In addition to this, we included only reports that were published in English or had verified English translations available, to avoid the potential for errors that may arise due to mistranslations by the current authors. Although this only resulted in the exclusion of two reports, it may introduce mono-language bias, which also limits generalizability. Our results therefore cannot be generalized beyond Western samples and only represent an analysis of the results of English-language journals.

General Conclusions

The results of the meta-analyses we have presented suggest that moral licensing, positive consistency, and moral compensation may represent real effects, but these effects occur in narrow conditions and are inflated by file-drawer bias. We found evidence for moral licensing only for studies that used negative target behaviors, positive consistency only for prosocial-FITD studies, and moral compensation only for studies using positive target behaviors. The pattern of results cannot be readily explained by current theoretical explanations of SMB effects, so further empirical research is needed to understand the mechanisms underlying SMB effects. The results of our analyses provide insight into the conditions under which different types of SMB effects occur and the approximate magnitude of these effects. Researchers in this field can use these insights to inform appropriate methodological decisions when designing future research. A more comprehensive understanding of when and how a person's moral decision making can be impacted by their previous moral behavior could assist in the development of strategies to increase positive behavior (and avoid triggering negative behavior) in the real world.

References

References included in Meta-Analysis 1 are marked with a single asterisk. References included in Meta-Analysis 2 are marked with a double asterisk.

**Abrahams, M. F., & Bell, R. A. (1994). Encouraging charitable contributions: An examination of three models of Door-in-the-Face compliance. *Communication Research*, *21*(2), 131–153. https://doi.org/10.1177/009365094021002001

*Aquino, K., Freeman, D., Reed, A., Lim, V. K., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, *97*(1), 123–141. https://doi.org/10.1037/a0015406

*Arnold, O., & Kaiser, F. G. (2016). Understanding the foot-in-the-door effect as a pseudo-effect from the perspective of the Campbell paradigm. *International Journal of Psychology*, *53*(2), 157–165. https://doi.org/10.1002/ijop.12289

Beaman, A. L., Cole, C. M., Preston, M., Klentz, B., & Steblay, N. M. (1983). Fifteen years of foot-in-the-door research: A meta-analysis. *Personality and Social Psychology Bulletin*, *9*(2), 181–196. https://doi.org/10.1177/0146167283092002

Bem, D. J. (1972). Self-perception theory. In L. Berkowiitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6

*Berger, L., Guo, L., & King, T. (2020). Selfish sharing? The impact of the sharing economy on tax reporting honesty. *Journal of Business Ethics*, *137*(2), 181–205. https://doi.org/10.1007/s10551-019-04409-z

*Bergquist, M. (2020). Most people think they are more pro-environmental than others: A demonstration of the better-than-average effect in perceived pro-environmental behavioral engagement. *Basic and Applied Social Psychology*, *42*(1), 50–61. https://doi.org/10.1080/01973533.2019.1689364

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

*Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2012). *Raw data* [Unpublished raw data].

*,**Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology*, *45*(3), 232–238. https://doi.org/10.1027/1864-9335/a000189

Borenstein, M., Hedges, L. V., Higgins, E. T., & Rothstein, H. R. (2014). *Comprehensive meta-analysis* (Version 3.0) [Computer software]. Biostat. https://meta-analysis.com/

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

**Boster, F. J., Cruz, S., Manata, B., DeAngelis, B. N., & Zhuang, J. (2016). A meta-analytic review of the effect on guilt and compliance. *Social Influence*, *11*(1), 54–67. https://doi.org/10.1080/15534510.2016.1142892

Bostyn, D. H., & Roets, A. (2016). The morality of action: The asymmetry between judgments of praise and blame in the action–omission effect. *Journal of Experimental Social Psychology*, *63*, 19–25. https://doi.org/10.1016/j.jesp.2015.11.005

*Braun, J., & Gollwitzer, M. (2012). Leniency for out-group offenders. *European Journal of Social Psychology*, *42*(7), 883–892. https://doi.org/10.1002/ejsp.1908

*Brown, R. P., Tamborski, M., Wang, X. Q., Barnes, C. D., Mumford, M. D., Connelly, S., & Devenport, L. D. (2011). Moral credentialing and the rationalization of misconduct *Ethics and Behavior*, *21*(1), 1–12. https://doi.org/10.1080/10508422.2011.537566

Burger, J. M. (1999). The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review*, *3*(4), 303–325. https://doi.org/10.1037/h0023552

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. https://doi.org/10.1038/nrn3475

*,**Cann, A., Sherman, S. J., & Elkes, R. (1975). Effects of initial request size and timing of a second request on compliance: The foot-in-the-door and the door in the face. *Journal of Personality and Social Psychology*, *32*(5), 774–782. https://doi.org/10.1037/0022-3514.32.5.774

Carlsmith, J. M., & Gross, A. E. (1969). Some effects of guilt on compliance. *Journal of Personality and Social Psychology*, *11*(3), 232–239. https://doi.org/10.1037/h0027039

*Carrico, A. R., Raimi, K. T., Truelove, H. B., & Eby, B. (2018). Putting your money where your mouth is: An experimental test of pro-environmental spillover from reducing meat consumption to monetary donations. *Environment and Behavior*, *50*(7), 723–748. https://doi.org/10.1177/0013916517713067

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

*Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating being good later allow you to be bad now? *Journal of Experimental Social Psychology*, *56*, 110–116. https://doi.org/10.1016/j.jesp.2014.09.009

*Chartrand, T., Pinckert, S., & Burger, J. M. (1999). When manipulation backfires: The effects of time delay and requester on the foot-in-the-door technique. *Journal of Applied Social Psychology*, *29*(1), 211–221. https://doi.org/10.1111/j.1559-1816.1999.tb01382.x

*Choi, B., Crandall, C. S., & La, S. (2014). Permission to be prejudiced: Legitimacy credits in the evaluation of advertisements. *Journal of Applied Social Psychology*, *44*(3), 190–200. https://doi.org/10.1111/jasp.12217

**Cialdini, R. B., Darby, B. L., & Vincent, J. E. (1973). Transgression and altruism: A case for hedonism. *Journal of Experimental Social Psychology*, *9*(6), 502–516. https://doi.org/10.1016/0022-1031(73)90031-0

**Cialdini, R. B., Vincent, J. E., Lewis, S. K., Catalan, J., Wheeler, D., & Darby, B. L. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology*, *31*(2), 206–215. https://doi.org/10.1037/h0076284

*Clot, S., Grolleau, G., & Ibanez, A. (2018). Moral self-licencing and social dilemmas: An experimental analysis from a taking game in Madagascar. *Applied Economics*, *50*(27), 2980–2991. https://doi.org/10.1080/00036846.2017.1412083

*Clot, S., Grolleau, G., & Ibanez, L. (2013). Self-licensing and financial rewards: Is morality for sale? *Economics Bulletin*, *33*(3), 2298–2306. http://EconPapers.repec.org/RePEc:ebl:ecbull:eb-13-00385

*Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters*, *123*(2), 191–194. https://doi.org/10.1016/j.econlet.2014.01.039

Coburn, K. M., & Vevea, J. L. (2017). *Weightr: Estimating weight-function models for publication bias* (R package for Version 1.1.2) [Compute software]. https://CRAN.R-project.org/package=weightr

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.

*Comello, M. L. G., Myrick, J. G., & Raphiou, A. L. (2016). A health fundraising experiment using the "Foot-in-the-Door" technique. *Health Marketing Quarterly*, *33*(3), 206–220. https://doi.org/10.1080/07359683.2016.1199209

*Conway, P., Cheung, I., & Olson, J. M. (2014). *Let he who is without sin cast the first stone: Activating the moral self amplifies moral judgments and behavior* [Manuscript under review].

*Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, *38*(7), 907–919. https://doi.org/10.1177/0146167212442394

*,**Cornelissen, G., Bashshur, M. R., Rode, J., & Le Menestrel, M. (2013). Rules or consequences? The role of ethical mind-sets in moral dynamics. *Psychological Science*, *24*(4), 482–488. https://doi.org/10.1177/0956797612457376

*Cornelissen, G., Dewitte, S., Warlop, L., & Yzerbyt, V. (2007). Whatever people say I am, that's what I am: Social labeling as a social marketing tool. *International Journal of Research in Marketing*, *24*(4), 278–288. https://doi.org/10.1016/j.ijresmar.2007.05.001

*Cornelissen, G., Pandelaere, M., Warlop, L., & Dewitte, S. (2008). Positive cueing: Promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing*, *25*(1), 46–55. https://doi.org/10.1016/j.ijresmar.2007.06.002

**Cryder, C. E., Springer, S., & Morewedge, C. K. (2012). Guilty feelings, targeted actions. *Personality and Social Psychology Bulletin*, *38*(5), 607–618. https://doi.org/10.1177/0146167211435796

**Cunningham, M. R., Steinberg, J., & Grev, R. (1980). Wanting to and having to help: Separate motivations for positive mood and guilt-induced helping. *Journal of Personality and Social Psychology*, *38*(2), 181–192. https://doi.org/10.1037//0022-3514.38.2.181

**Darlington, R., & Macker, C. (1966). Displacement of guilt-produced altruistic behavior. *Journal of Personality and Social Psychology*, *4*(4), 442–443. https://doi.org/10.1037/h0023743

**de Hooge, I. E. (2013). Moral emotions and prosocial behavior: It may be time to change our view of shame and guilt. In C. Mohiyeddini, M. Eysenck, S. Bauer, C. Mohiyeddini, M. Eysenck, & S. Bauer (Eds.), *Handbook of psychology of emotions: Recent theoretical perspectives and novel empirical findings* (2nd ed., pp. 255–275). Nova Science Publishers.

**de Hooge, I. E., Breugelmans, S. M., & Zeelenberg, M. (2008). Not so ugly after all: When shame acts as a commitment device. *Journal of Personality and Social Psychology*, *95*(4), 933–943. https://doi.org/10.1037/a0011991

**de Hooge, I. E., Nelissen, R. M. A., Breugelmans, S. M., & Zeelenberg, M. (2011). What is moral about guilt? Acting "prosocially" at the disadvantage of others. *Journal of Personality and Social Psychology*, *100*(3), 462–473. https://doi.org/10.1037/a0021459

DeJong, W. (1979). An examination of self-perception mediation of the foot-in-the-door effect. *Journal of Personality and Social Psychology*, *37*(12), 2221–2239. https://doi.org/10.1207/s15327957pspr0304_2

Dillard, J. P., Hunter, J. E., & Burgoon, M. (1984). Sequential-request persuasive strategies: A meta-analysis of foot-in-the-door and door-in-the-face. *Human Communication Research*, *10*(4), 461–488. https://doi.org/10.1111/j.1468-2958.1984.tb00028.x

**Ding, W., Xie, R., Sun, B., Li, W., Wang, D., & Zhen, R. (2016). Why does the "sinner" act prosocially? The mediating role of guilt and the moderating role of moral identity in motivating moral cleansing. *Frontiers in Psychology*, *7*, Article 1317. https://doi.org/10.3389/fpsyg.2016.01317

*Dolinski, D. (2012). The nature of the first small request as a decisive factor in the effectiveness of the foot-in-the-door technique. *Applied Psychology*, *61*(3), 437–454. https://doi.org/10.1111/j.1464-0597.2011.00477.x

**Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology*, *27*(1), 94–100. https://doi.org/10.1037/h0035582

*Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

*Eby, B., Carrico, A. R., & Truelove, H. B. (2019). The influence of environmental identity labeling on the uptake of pro-environmental behaviors. *Climatic Change*, *155*(4), 563–580. https://doi.org/10.1007/s10584-019-02501-x

*Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology*, *45*(3), 590–593. https://doi.org/10.1016/j.jesp.2009.02.001

Effron, D. A., & Conway, P. (2015). When virtue leads to villainy: Advances in research on moral self-licensing. *Current Opinions in Psychology*, *6*, 32–35. https://doi.org/10.1016/j.copsyc.2015.03.017

*Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology*, *103*(6), 916–932. https://doi.org/10.1037/a0030008

Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions? *Personality and Social Psychology Bulletin*, *36*(12), 1618–1634. https://doi.org/10.1177/0146167210385922

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, *79*(6), 861–875. https://doi.org/10.1037//0022-3514.79.6.861

**Even-Chen, M., Yinon, Y., & Bizman, A. (1978). The door in the face technique: Effects of the size of the initial request. *European Journal of Social Psychology*, *8*(1), 135–140. https://doi.org/10.1002/ejsp.2420080113

*Falomir-Pichastor, J. M., Mugnry, G., Frederic, N., Berent, J., & Lalot, F. (2018). Motivation to maintain a nonprejudiced self-identity: The role of justification for prejudice and normative context on moral self-licensing. *Social Psychology*, *49*(3), 168–181. https://doi.org/10.1027/1864-9335/a000339

*Fanghella, V., d'Adda, G., & Tavoni, M. (2019). On the use of nudges to affect spillovers in environmental behaviors. *Frontiers in Psychology*, *10*, Article 61. https://doi.org/10.3389/fpsyg.2019.00061

*,**Fanghella, V., & Thogersen, J. (2022). Experimental evidence of moral cleansing in the interpersonal and environmental domains. *Journal of Behavioral and Experimental Economics*, *97*, Article 101838. https://doi.org/10.1016/j.socec.2022.101838

Feeley, T. H., Anker, A. E., & Aloe, A. M. (2012). The door-in-the-face persuasive message strategy: A meta-analysis of the first 35 years. *Communication Monographs*, *79*(3), 316–343. https://doi.org/10.1080/03637751.2012.697631

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001

*,**Ferguson, R. (2019). *How flexible is morality? A test of the moral credits model of moral balancing* [Doctoral dissertation, Australian Catholic University]. Australian Catholic University Research Bank. https://web.archive.org/web/20200315063330id_/https://researchbank.acu.edu.au/cgi/viewcontent.cgi?article=1799&context=theses

Ferguson, R., Kaufmann, L., Brown, A., & de la Piedad Garcia, X. (2023). *Sequential moral behaviour meta-analyses (moral balancing and moral consistency)*. Open Science Framework. https://osf.io/upby2

Fern, E. F., Monroe, K. B., & Avila, R. A. (1986). Effectiveness of multiple request strategies: A synthesis of research results. *Journal of Marketing Research*, *23*(2), 144–152. https://doi.org/10.2307/3151661

Fisher, Z., & Tipton, E. (2015). *robumeta: An R-package for robust variance estimation in meta-analysis*. arXiv. https://doi.org/10.48550/arXiv.1503.02220

*Fointiat, V. (2006). "You're helpful" versus "that's clear." Social versus functional label in the foot-in-the-door paradigm. *Social Behavior and Personality*, *34*(5), 461–465. https://doi.org/10.1037/a0021459

*Fointiat, V., & Saint-Bauzel, R. (2010). Appearance (ethnic origin) of requester affects the foot-in-the-door technique. *Perceptual and Motor Skills*, *111*(3), 694–700. https://doi.org/10.2466/07.17.PMS.111.6.694-700

*Foss, R. D., & Dempsey, C. B. (1979). Blood donation and the foot-in-the-door technique: A limiting case. *Journal of Personality and Social Psychology*, *37*(4), 580–590. https://doi.org/10.1037//0022-3514.37.4.580

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195–202. https://doi.org/10.1037/h0023552

**Freedman, J. L., Wallington, S. A., & Bless, E. (1967). Compliance without pressure: The effect of guilt. *Journal of Personality and Social Psychology*, *7*(2, Pt. 1), 117–124. https://doi.org/10.1037/h0025009

*Frimer, J., Decter-Frain, A., & Zhu, L. (2015). *Do givers leak their social motives through what they talk about? Maybe not* [Unpublished manuscript].

*Gamian-Wilk, M., & Lachowicz-Tabaczek, K. (2007). Implicit theories and compliance with the foot-in-the-door technique. *Polish Psychological Bulletin*, *38*, 50–63. https://www.researchgate.net/publication/306082565_Implicit_theories_and_compliance_with_the_foot-in-the-door_technique

*,**Gamian-Wilk, M., & Lachowicz-Tabaczek, K. (2009). The belief to have fixed or malleable traits and help giving: Implicit theories and sequential social influence techniques. *Polish Psychological Bulletin*, *40*(2), 85–100. https://doi.org/10.2478/s10059-009-0009-z

*Garvey, A. M., & Bolton, L. E. (2017). Eco-product choice cuts both ways: How pro-environmental licensing versus reinforcement is contingent upon environmental consciousness. *Journal of Public Policy and Marketing*, *36*(2), 284–298. https://doi.org/10.1509/jppm.16.096

Geiger, S. J., Brick, C., Nalborczyk, L., Bosshard, A., & Jostmann, N. B. (2021). More green than gray? Toward a sustainable overview of environmental spillover effects: A Bayesian meta-analysis. *Journal of Environmental Psychology*, *78*, Article 101694. https://doi.org/10.1016/j.jenvp.2021.101694

*Geng, L., Cheng, X., Tang, Z., Zhou, K., & Ye, L. (2016). Can previous pro-environmental behaviours influence subsequent environmental behaviours? The licensing effect of pro-environmental behaviours. *Journal of Pacific Rim Psychology*, *10*, 1–9. https://doi.org/10.1017/prp.2016.6

**Genschow, O., Westfal, M., Crusius, J., Bartosch, L., Feikes, K. I., Pallasch, N., & Wozniak, M. (2021). Does social psychology persist over half a century? A direct replication of Cialdini et al.'s (1975) classic door-in-the-face technique. *Journal of Personality and Social Psychology*, *120*(2), e1–e7. https://doi.org/10.1037/pspa0000261

*Ghesla, C., Grieder, M., & Schmitz, J. (2019). Nudge for good? Choice defaults and spillover effects. *Frontiers in Psychology*, *10*, Article 178. https://doi.org/10.3389/fpsyg.2019.00178

**Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological Science*, *21*(5), 712–720. https://doi.org/10.1177/0956797610366545

*Girandola, F. (2002). Sequential requests and organ donation. *The Journal of Social Psychology*, *142*(2), 171–178. https://doi.org/10.1080/00224540209603893

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behavior. *Management Science*, *58*(1), 179–187. https://doi.org/10.1287/mnsc.1110.1437

**Goldman, M., McVeigh, J. F., & Richterkessing, J. L. (1984). Door-in-the-face procedure: Reciprocal concession, perceptual contrast, or worthy person. *The Journal of Social Psychology*, *123*(2), 245–251.

*Goldman, M., Seever, M., & Seever, M. (1982). Social labeling and the foot-in-the-door effect. *The Journal of Social Psychology*, *117*(1), 19–23. https://doi.org/10.1080/00224545.1982.9713402

*,**Greene, M., & Low, K. (2014). Public integrity, private hypocrisy, and the moral licensing effect. *Social Behavior and Personality*, *42*(3), 391–400. https://doi.org/10.2224/sbp.2014.42.3.391

*Guadagno, R. E., Asher, T., Demaine, J., & Cialdini, R. B. (2001). When saying yes leads to saying no: Preference for consistency and the reverse foot-in-the-door effect. *Personality and Social Psychology Bulletin*, *27*(7), 859–867. https://doi.org/10.1177/0146167201277008

**Guéguen, N. (2003). Fund-raising on the web: The effect of an electronic door-in-the-face technique on compliance to a request. *CyberPsychology & Behavior*, *6*(2), 189–193. https://doi.org/10.1089/109493103321640383

**Guéguen, N. (2014). Door-in-the-face technique and delay to fulfill the final request: An evaluation with a request to give blood. *The Journal of Psychology*, *148*(5), 569–576. https://doi.org/10.1080/00223980.2013.817963

*Guéguen, N., & Pascual, A. (2015). Foot-in-the-door technique and problematic implicit request for help. *Swiss Journal of Psychology*, *74*(2), 111–114. https://doi.org/10.1024/1421-0185/a000155

Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology*, *26*(2), 201–218. https://doi.org/10.1002/(SICI)1099-0992(199603)26:2<201::AID-EJSP745>3.0.CO;2-J

Hansen, R. A., & Robinson, L. M. (1980). Testing the effectiveness of alternative foot-in-the-door manipulations. *Journal of Market Research*, *17*(3), 359–364. https://doi.org/10.1177/002224378001700309

*Harris, M. B. (1972). The effects of performing one altruistic act on the likelihood of performing another. *The Journal of Social Psychology*, *88*(1), 65–73. https://doi.org/10.1080/00224545.1972.9922544

*,**Hayley, A., & Zinkiewicz, L. (2013). Does moral cleansing moderate the effect of evolutionary altruism on helping intention? An exploratory study. *Journal of Social, Evolutionary, and Cultural Psychology*, *7*(1), 24–35. https://doi.org/10.1037/h0099176

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

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

**Henderson, M. D., & Burgoon, E. M. (2013). Why the door-in-the-face technique can sometimes backfire: A construal-level account. *Social Psychological and Personality Science*, *5*(4), 475–483. https://doi.org/10.1177/1948550613506719

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2–3), 111–135. https://doi.org/10.1017/S0140525X10000725

Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior? A meta-analysis. *Review of General Psychology*, *20*(2), 129–140. https://doi.org/10.1037/gpr0000062

*,**Ho, B., Taber, J., Poe, G., & Bento, A. (2015). The effects of moral licensing and moral cleansing in contingent valuation and laboratory experiments on the demand to reduce externalities. *Environmental and Resource Economics*, *64*(2), 317–340. https://doi.org/10.1007/s10640-014-9872-y

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality of everyday life. *Science*, *348*(6202), 1340–1343. https://doi.org/10.1126/science.aac5401

**Hu, M. L. M., Horng, J. S., Teng, C. C., Chiou, W. B., & Yen, C. D. (2014). Fueling green dining intention: The self-completion theory perspective. *Asia Pacific Journal of Tourism Research*, *19*(7), 793–808. https://doi.org/10.1080/10941665.2013.806941

Janoff-Bulman, R. (2012). Conscience: The dos and don'ts of moral regulation. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 131–148). American Psychological Association. https://doi.org/10.1037/13091-007

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*(3), 521–537. https://doi.org/10.1037/a0013779

*Jasinskaja-Lahti, I., Vezzali, L., Ranta, M., Pacilli, M. G., Giacomantonio, M., & Pagliaro, S. (2021). Conditional secondary transfer effect: The moderating role of moral credentials and prejudice. *Group Processes & Intergroup Relations*, *24*(7), 1219–1237. https://doi.org/10.1177/1368430220940401

*Jeong, H. J., & Koo, D. M. (2015). Volunteering as a mechanism to reduce guilt over purchasing luxury items. *Journal of Product and Brand Management*, *24*(7), 758–769. https://doi.org/10.1108/jpbm-01-2015-0784

*Jones, J. A., & Koenig, A. M. (2018). The effect of "socially moral" purchases on future financial donations. *International Journal of Nonprofit and Voluntary Sector Marketing*, *23*(4), Article e1614. https://doi.org/10.1002/nvsm.1614

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The moral self-image scale: Measuring and understanding the malleability of the moral self. *Frontiers in Psychology*, *6*, Article 1878. https://doi.org/10.3389/fpsyg.2015.01878

*,**Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality & Social Psychology Bulletin*, *37*(5), 701–713. https://doi.org/10.1177/0146167211400208

**Kahwaty, R. C. (1979). *Helping behavior as a consequence of undiscovered and discovered transgressions* [Doctoral dissertation, City University of New York]. Dissertation Abstracts International Section A: Humanities and Social Sciences.

*Karmarker, U., & Bollinger, B. (2015). BYOB: How bringing your own shopping bags leads to treating yourself and the environment. *Journal of Marketing*, *79*(4), 1–15. https://doi.org/10.1509/jm.13.0228

*Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, *43*(2), 259–266. https://doi.org/10.1509/jmkr.43.2.259

**Konecni, V. J. (1972). Some effects of guilt on compliance: A field replication. *Journal of Personality and Social Psychology*, *23*(1), 30–32. https://doi.org/10.1037/h0032875

**Konoske, P., Staple, S., & Graf, R. G. (1979). Compliant reactions to guilt: Self-esteem or self-punishment. *Journal of Social Psychology*, *108*(2), 207–211. https://doi.org/10.1080/00224545.1979.9711633

Kordes-de Vaal, J. (1996). Intention and the omission bias: Omissions perceived as nondecisions. *Acta Psychologica*, *93*(1–3), 161–172. https://doi.org/10.1016/0001-6918(96)00027-3

*Kristofferson, K., White, K., & Peloza, J. (2014). The nature of slacktivism: How the social observability of an initial act of token support affects subsequent prosocial action. *Journal of Consumer Research*, *40*(6), 1149–1166. https://doi.org/10.1086/674137

Kuper, N., & Bott, A. (2019). Has the evidence for moral licensing been inflated by publication bias? *Meta-Psychology*, *3*, 1–22. https://doi.org/10.15626/MP.2018.878

Lasarov, W., & Hoffmann, S. (2020). Social moral licensing. *Journal of Business Ethics*, *165*, 45–66. https://doi.org/10.1007/s10551-018-4083-z

**Lee, C., Hochman, G., Prince, H. E., & Ariely, D. (2016). Past actions as self-signals: How acting in a self-interested way influences environmental decision making. *PLOS ONE*, *11*(7), Article e0158456. https://doi.org/10.1371/journal.pone.0158456

*Li, A., Mai, K. M., & Bagger, J. (2017). Licensed to say no: How and why does engaging in a prior moral action influence family support provision? *Journal of Vocational Behavior*, *102*, 86–98. https://doi.org/10.1016/j.jvb.2017.07.005

*Lin, Y., & Chang, C. A. (2017). Exploring wasteful consumption. *Journal of Environmental Psychology*, *49*, 106–111. https://doi.org/10.1016/j.jenvp.2017.01.001

*Longoni, C., Gollwitzer, P. M., & Oettingen, G. (2014). A green paradox: Validating green choices has ironic effects on behavior, cognition, and perception. *Journal of Experimental Social Psychology*, *50*, 158–165. https://doi.org/10.1016/j.jesp.2013.09.010

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

*Maki, A. (2015). *The spread of behavior: When, how, and for whom do proenvironmental behaviors spread to other people and other behaviors?* [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations and Theses Global. https://www.proquest.com/dissertations-theses/spread-behavior-when-how-whom-do-proenvironmental/docview/1733999976/se-2

*Mann, N. H., & Kawakami, K. (2011). The long, steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General*, *141*(1), 187–197. https://doi.org/10.1037/a0025602

*Mazar, N., & Zhong, C. (2010). Do green products make us better people? *Psychological Science*, *21*(4), 494–498. https://doi.org/10.1177/0956797610363538

**McMillen, D. L. (1971). Transgression, self-image, and compliant behavior. *Journal of Personality and Social Psychology*, *20*(2), 176–179. https://doi.org/10.1037/h0031686

**McMillen, D. L., Jackson, J. A., & Austin, J. B. (1974). Effects of positive and negative requests on compliance following transgression. *Bulletin of the Psychonomic Society*, *3*(1B), 80–82. https://doi.org/10.3758/BF03333403

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

**Meineri, S., Dupré, M., Guéguen, N., & Vallée, B. (2016). Door-in-the-face and but-you-are-free: Testing the effect of combining two no-pressure compliance paradigms. *Psychological Reports*, *119*(1), 276–289. https://doi.org/10.1177/0033294116657064

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, *4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

**Millar, M. G. (2002a). Effects of a guilt induction and guilt reduction on door-in-the-face. *Communication Research*, *29*(6), 666–680. https://doi.org/10.1177/009365002237831

**Millar, M. G. (2002b). The effectiveness of the door-in-the-face compliance strategy on friends and strangers. *The Journal of Social Psychology*, *142*(3), 295–304. https://doi.org/10.1080/00224540209603901

Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43, pp. 115–155). Academic Press. https://doi.org/10.1016/S0065-2601(10)43003-8

Miller, R. L., Seligman, C., Clark, N. T., & Bush, M. (1976). Perceptual contrast versus reciprocal concession as mediators of induced compliance. *Canadian Journal of Behavioral Science*, *8*(4), 401–409. https://doi.org/10.1037/h0081965

*Miller, R. L., & Suls, J. (1977). Helping, self-attribution, and the size of an initial request. *Journal of Social Psychology*, *103*(2), 203–208. https://doi.org/10.1080/00224545.1977.9713318

*Millet, K., Du, G., Cabooter, E., & Weijters, B. (2022). The limited impact of positive cueing on pro-environmental choices. *Journal of Environmental Psychology*, *79*, Article 101732. https://doi.org/10.1016/j.jenvp.2021.101732

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In D. Narvaez & D. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 341–354). Cambridge University Press.

*Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*(1), 33–43. https://doi.org/10.1037/0022-3514.81.1.33

**Mueb, L., Proeger, T., Schneider, T., & Bizer, K. (2016). The victim matters—Experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters*, *23*(16), 1162–1167. https://doi.org/10.1080/13504851.2016.1139674

*Mueller, A. B., & Skitka, L. J. (2015). *Licensed to lie: Evidence for moral licensing in persuasive communication contexts* [Unpublished raw data].

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology*, *67*, 363–385. https://doi.org/10.1146/annurev-psych-010213-115120

*,**Nagel, V. (2014). *Three essays on moral self-regulation of honesty and impression management* [Doctorial dissertation, University of Passau]. Dissertation Abstracts International Section A: Humanities and Social Sciences.

**Neufeld, S. L. (2012). *Self-conscious cooperation: Implications of a functional approach to emotions for behavior in social dilemmas* [Doctoral dissertation, Arizona State University]. Dissertation Abstracts International Section A: Humanities and Social Sciences.

Nisan, M. (1990). Moral balance: A model of how people arrive at moral decisions. In T. E. Wren (Ed.), *The moral domain: Essays in the ongoing discussion between philosophy and the social sciences* (pp. 283–314). MIT Press.

Nisan, M. (1991). The moral balance model: Theory and research extending our understanding of moral choice and deviation. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 3, pp. 213–249). Lawrence Erlbaum.

*Noblet, C. L., & McCoy, S. K. (2018). Does one good turn deserve another? Evidence of domain-specific licensing in energy behavior. *Environment and Behavior*, *50*(8), 839–863. https://doi.org/10.1177/0013916517718022

**O'Keefe, D. J., & Figgé, M. (1997). A guilt-based explanation of the door-in-the-face influence strategy. *Human Communication Research*, *24*(1), 64–81. https://doi.org/10.1111/j.1468-2958.1997.tb00587.x

O'Keefe, D. J., & Hale, S. L. (1998). The door-in-the-face influence strategy: A random-effects meta-analytic review. *Communication Yearbook*, *21*(1), 1–33. https://doi.org/10.1080/23808985.1998.11678947

Palmer, T. M., Peters, J. L., Sutton, A. J., & Moreno, S. G. (2008). Contour-enhanced funnel plots for meta-analysis. *Stata Journal*, *8*(2), 242–254. https://doi.org/10.1177/1536867X0800800206

**Pascual, A., Dagot, L., Vallee, B., & Gueguen, N. (2009). Compliance without pressure, mediatization of a Tsunami and charitable donation: Compared effectiveness of the door-in-the-face and "you are free of …" techniques. *European Review of Applied Psychology*, *59*(1), 79–84. https://doi.org/10.1016/j.erap.2006.07.003

**Pascual, A., & Gueguen, N. (2006). Door-in-the-face technique and monetary solicitation: An evaluation in a field setting. *Perceptual and Motor Skills*, *103*(3), 974–978. https://doi.org/10.2466/pms.103.3.974-978

*Pascual, A., Gueguen, N., Pujos, S., & Felonneau, M. L. (2013). Foot-in-the-door and problematic requests: A field experiment. *Social Influence*, *8*(1), 46–53. https://doi.org/10.1080/15534510.2012.696038

Pascual, A., & Guéguen, N. (2005). Foot-in-the-door and door-in-the-face: A comparative meta-analytic study. *Psychological Reports*, *96*(1), 122–128. https://doi.org/10.2466/pr0.96.1.122-128

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2011). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Pendleton, M. G., & Batson, C. D. (1979). Self-presentation and the door-in-the-face technique for inducing compliance. *Personality and Social Psychology Bulletin*, *5*(1), 77–81. https://doi.org/10.1177/014616727900500117

**Peters, J. E. (1974). *The effects of locus-of-control and temporary affective states of the benefactor on altruism* [Doctoral dissertation, University of Montana]. ProQuest Dissertations and Theses Global. https://www.proquest.com/openview/71acbb1eb1e0d9b01d1e30bd8e31e266/1?pq-origsite=gscholar&cbl=18750&diss=y

**Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, *93*, 374–383. https://doi.org/10.1016/j.jebo.2013.03.030

*Raska, D. (2010). *Licensing and fluency of sacrosanct experience recall* [PhD thesis, Washington State University]. Dissertation Abstracts International Section A: Humanities and Social Sciences.

**Reeves, R. A., Baker, G. A., Boyd, J. G., & Cialdini, R. B. (1991). The door-in-the-face technique: Reciprocal concessions vs. self-presentational explanations. *Journal of Social Behavior and Personality*, *6*(3), 545–558.

**Regan, J. W. (1971). Guilt, perceived injustice, and altruistic behavior. *Journal of Personality and Social Psychology*, *18*(1), 124–132. https://doi.org/10.1037/h0030712

*,**Reingen, P. H. (1978). On inducing compliance with requests. *Journal of Consumer Research*, *5*(2), 96–102. https://doi.org/10.1086/208720

*Rind, B., & Benjamin, D. (1994). Effects of public image concerns and self-image on compliance. *Journal of Social Psychology*, *134*(1), 19–25. https://doi.org/10.1080/00224545.1994.9710878

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

*,**Rotella, A., & Barclay, P. (2020). Failure to replicate moral licensing and moral cleansing in an online experiment. *Personality and Individual Differences*, *161*, Article 109967. https://doi.org/10.1016/j.paid.2020.109967

**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

*,**Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*(4), 523–528. https://doi.org/10.1111/j.1467-9280.2009.02326.x

*Saint-Bauzel, R., & Fointiat, V. (2012). The sweet smell of the requester: Vanilla, camphor, and foot-in-the-door. *Social Behavior and Personality*, *40*(3), 369–374. https://doi.org/10.2224/sbp.2012.40.3.369

Salomon, E., Preston, J. L., & Tannenbaum, M. B. (2017). Climate change helplessness and the (de)moralization of individual energy behavior. *Journal of Experimental Psychology: Applied*, *23*(1), 15–28. https://doi.org/10.1037/xap0000105

*,**Schlegelmilch, B. B., & Simbrunner, P. (2019). Moral licensing and moral cleansing applied to company-NGO collaborations in an online context. *Journal of Business Research*, *95*, 544–552. https://doi.org/10.1016/j.jbusres.2018.07.040

*,**Schwabe, M., Dose, D. B., & Walsh, G. (2017). Every saint has a past, and every sinner has a future: Influences of regulatory focus on consumers' moral self-regulation. *Journal of Consumer Psychology*, *28*(2), 234–252. https://doi.org/10.1002/jcpy.1025

Seligman, C., Bush, M., & Kirsch, K. (1976). Relationship between compliance in the foot-in-the-door paradigm and size of first request. *Journal of Personality and Social Psychology*, *33*(5), 517–520. https://doi.org/10.1037/0022-3514.33.5.517

*Shaw, M. P. (2010). *Prejudice with a conscience: How a strong moral identity relates to greater prejudice* [PhD thesis, University of Texas at El Paso]. ProQuest Dissertations and Theses Global.

Simbrunner, P., & Schlegelmilch, B. B. (2017). Moral licensing: A culture-moderated meta-analysis. *Management Review Quarterly*, *67*(4), 201–225. https://doi.org/10.1007/s11301-017-0128-0

*Skandrani-Marzouki, I., Marzouki, Y., & Joule, R. (2012). Effects of subliminal affective priming on helping behavior using the foot-in-the-door technique. *Psychological Reports*, *111*(3), 669–686. https://doi.org/10.2466/07.17.21.PR0.111.6.669-686

*Smith, S. R. (2020). *Moderators of positive and negative spillover* [Master's thesis, University of North Florida]. UNF Digital Commons. https://digitalcommons.unf.edu/etd/889

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–101. https://doi.org/10.1016/0022-1031(91)90011-T

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

*Stellar, J. E., & Willer, R. (2014). The corruption of value: Negative moral associations diminish the value of money. *Social Psychological and Personality Science*, *5*(1), 60–66. https://doi.org/10.1177/1948550613484770

*Susewind, M., & Hoelzl, E. (2014). A matter of perspective: Why past moral behavior can sometimes encourage and other times discourage future moral striving. *Journal of Applied Social Psychology*, *44*(3), 201–209. https://doi.org/10.1111/jasp.12214

Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are we more moral than we think? Exploring the role of affect in moral behavior and moral forecasting. *Psychological Science*, *22*(4), 553–558. https://doi.org/10.1177/0956797611402513

Teper, R., Tullet, A. M., Page-Gould, E., & Inzlicht, M. (2015). Errors in moral forecasting: Perceptions of affect shape the gap between moral behaviors and moral forecasts. *Personality and Social Psychology Bulletin*, *41*(7), 887–900. https://doi.org/10.1177/0146167215583848

**Terrier, L., Joule, R., & Marfaing, B. (2011). Requester's acceptance and non-acceptance of the refusal of the initial request: How to improve the door-in-the-face effects? *Current Research in Social Psychology*, *17*, Article 1.

**Terrier, L., Marfaing, B., & Boldi, M. (2013). Door-in-the-face: Is it really necessary that both requests be made by the same requester? *Psychological Reports*, *113*(2), 675–682. https://doi.org/10.2466/17.07.PR0.113x25z7

Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, *57*, 160–171. https://doi.org/10.1016/j.enpol.2013.01.021

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. https://doi.org/10.1016/j.gloenvcha.2014.09.004

*,**Truelove, H. B., Yeung, K. L., Carrico, A. R., Gillis, A. J., & Raimi, K. T. (2016). From plastic bottle recycling to policy support: An experimental test of pro-environmental spillover. *Journal of Environmental Psychology*, *46*, 55–66. https://doi.org/10.1016/j.jenvp.2016.03.004

**Turner, M. M., Tamborini, R., Limon, M. S., & Zuckerman-Hyman, C. (2007). The moderators and mediators of door-in-the-face requests: Is it a negotiation or a helping experience? *Communication Monographs*, *74*(3), 333–356. https://doi.org/10.1080/03637750701543469

*Urban, J., Bahník, Š., & Kohlová, M. B. (2019). Green consumption does not make people cheat: Three attempts to replicate moral licensing effect due to pro-environmental behavior. *Journal of Environmental Psychology*, *63*, 139–147. https://doi.org/10.1016/j.jenvp.2019.01.011

Van de Ven, N., Blanken, I., & Zeelenberg, M. (2018). Temptation-based reasoning: When tempted, everything becomes a (better) reason to indulge. *Journal of Marking Behavior*, *3*(3), 185–209. https://doi.org/10.1561/107.00000047

Viechtbauer, W. (2015). *Meta-analysis package for R: Package 'metafor'* (R package Version 1.9-5) [Computer software]. http://www.metafor-project.org/doku.php

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Villegas de Posada, C., & Vargas-Trujillo, E. (2015). Moral reasoning and personal behavior: A meta-analytical review. *Review of General Psychology*, *19*(4), 408–424. https://doi.org/10.1037/gpr0000053

**Ward, S. J., & King, L. A. (2018). Moral self-regulation, moral identity, and religiosity. *Personality and Individual Differences*, *115*(3), 495–525. https://doi.org/10.1037/pspp0000207

Wu, P., & Liu, H. (2014). Association between moral reasoning and moral behavior: A systematic review and meta-analysis. *Acta Psychologica Sinica*, *46*(8), 1192–1207. https://doi.org/10.3724/SP.J.1041.2014.01192

*Yam, K. C., Klotz, A. C., He, W., & Reynolds, S. J. (2017). From good soldiers to psychologically entitled: Examining when and why citizenship behavior leads to deviance. *Academy of Management Journal*, *60*(1), 373–396. https://doi.org/10.5465/amj.2014.0234

Yeung, S. K., Tay, T., & Feldman, G. (2022). Action and inaction in moral judgments and decisions: Meta-analysis of omission bias and omission-commission asymmetries. *Personality and Social Psychology Bulletin*, *48*(10), 1499–1515. https://doi.org/10.1177/01461672211042315

*,**Young, L., Chakroff, A., & Tom, J. (2012). Doing good leads to more good: The reinforcing power of a moral self-concept. *Review of Philosophy and Psychology*, *3*(3), 325–334. https://doi.org/10.1007/s13164-012-0111-6

*Zawadzki, S. J. (2015). *The interdependent nature of environmental behaviors: Testing a conceptual framework for behavioral spillover* [Master's thesis, Penn State University]. PennState University Libraries. https://etda.libraries.psu.edu/catalog/26452

*Zhang, Y., Jiang, J., Sun, Y., Gu, D., & Jiang, W. (2021). Engagement in cause-related marketing reduces pro-environmental behaviors. *Environment and Behavior*, *53*(10), 1047–1069. https://doi.org/10.1177/0013916520942602

Zhong, C., Ku, G. L., Lount, R. B., & Murnighan, J. K. (2010). Compensatory ethics. *Journal of Business Ethics*, *92*(3), 323–339. https://doi.org/10.1007/s10551-009-0161-6

Received June 27, 2021
Revision received March 29, 2024
Accepted April 11, 2024 ■