Apologies and moral behavior: How apologizing versus not apologizing influences moral licensing and moral cleansing

by

Justin Michael Ludwig

B.A., California State University, Fullerton, 2012

M.A., California State University, Fullerton, 2017

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Justin Michael Ludwig

It was defended on

October 20, 2022

and approved by

Kevin Binning, Ph.D, Associate Professor, Department of Psychology

Amanda Forest, Ph.D, Associate Professor, Department of Psychology

Thesis Advisor/Dissertation Director: Karina Schumann, Ph.D, Associate Professor, Department of Psychology

Apologies and moral behavior: How apologizing versus not apologizing influences moral licensing and moral cleansing

Justin Michael Ludwig, M.S.

University of Pittsburgh, 2022

Apologies are a tool for moral repair after an interpersonal offense. Yet, little is known about the extent to which apologies repair transgressors' moral self-image and how this might affect their subsequent behavior. Investigating these potential negative or positive downstream consequences of apologies offers insights into whether there are conditions under which apologies might backfire, freeing transgressors up to withdraw from subsequent opportunities to engage in moral behavior. In Study 1 (N = 243), using a hypothetical offense, I found that apologizing not only repaired moral self-image but also bolstered moral self-image above baseline levels. In Study 2 (N = 211), I examined whether transgressors who apologize (vs. do not apologize) to a victim feel licensed to engage in less moral subsequent behavior. Participants read the same offense from Study 1 and then imagined a situation where they could take credit for a coworker's ideas. Findings from Study 1 were replicated, however apologizing did not lead to moral licensing. Withholding an apology led to moral cleansing, but only under high levels of perceived negativity of the offense. In Study 3 (N = 593), participants recalled an unresolved conflict, wrote an apology letter (vs. no apology and control), and then had an opportunity to donate bonus money to a non-profit organization. Apologizing did not license participants to donate less bonus money and withholding an apology did not lead to moral cleansing. These studies suggest that apologizing has restorative effects on transgressors' moral self-image whereas withholding an apology could possibly leave transgressors in a morally depleted state. However, these studies did not consistently demonstrate

that apologizing vs. not apologizing impacted subsequent moral behavior in the form of moral licensing and moral cleansing. I discuss the limitations and potential reasons for these null findings, as well as directions for future research.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Decades of research have investigated the restorative power of an apology following an interpersonal offense (Darby & Schlenker, 1982; Exline et al., 2007; McCullough et al., 1997; Fehr et al., 2010; Kirchhoff et al., 2012; Schumann 2018; Schumann & Dragotta, 2021; Zechmeister et al., 2004). Apologies are strong predictors of victim forgiveness (Fehr et al., 2010), promoting relational repair in part by signaling the value that both transgressors and victims stake in their relationship with each other (Forster et al., 2021). Offering an apology can attenuate feelings of guilt and shame experienced by the transgressor and promote self-forgiveness (Fisher & Exline, 2010; Hall & Fincham, 2005; Witvliet et al., 2001), and receiving an apology can attenuate feelings of anger and hostility experienced by the victim and promote reconciliation (McCullough et al., 1997). The apology-forgiveness cycle has been examined using longitudinal designs, demonstrating their benefits over time (McCullough et al., 2014), and apologies might even help mitigate the insidious physiological effects that are associated with unresolved, hostile conflicts (Keicolt-Glaser & Newton, 2001; Witvliet et al., 2020).

Yet, despite the immediate and long-term benefits of apologies, little is known about how they impact transgressors' subsequent moral behavior. Although a growing body of research has identified important antecedents to transgressors' apologies (Schumann, 2018), almost no work to my knowledge has examined the downstream consequences of apologizing for transgressors' decision-making and behavior. In the current studies, I investigated how apologizing (vs. not apologizing) for an interpersonal offense affects (a) transgressors' moral self-image, and (b) subsequent moral behavior. If apologizing repairs (and possibly bolsters) transgressors' sense of moral self-image, apologizing might subsequently lead to moral licensing or moral consistency. It

is possible that transgressors who apologize (vs. do not apologize) to a victim might feel morally licensed to engage in less moral subsequent behavior, possibly even increasing their likelihood of offending another person. Alternatively, apologizing might signal identification with moral values and thus lead to moral consistency or subsequent prosocial behavior. Below, I review literature that supports the plausibility of both these hypotheses and describe my theoretical reasoning for predicting a moral licensing effect.

1.1 Apologies and Moral Repair

Transgressors commit interpersonal offenses when they violate the implicit or explicit values, rules, and norms shared between them and the victim (Woodyatt et al., 2022). According to the needs-based model of reconciliation, transgressors experience a threat to their moral self-image after committing an interpersonal offense (Schnabel & Nadler, 2008). In response to this threat, they are motivated to repair their moral self-image (Steele et al., 1993, Woodyatt et al., 2022). One way they might attempt to repair their moral self-image is by engaging in self-protection strategies such as defensiveness or taking no reparative actions (Woodyatt & Wenzel, 2013). Alternatively, they might engage in reparative actions such as apologizing (Cryder et al., 2012; Howell et al., 2012; Leunissen et al., 2017).

Whereas defensiveness (Woodyatt & Wenzel, 2013) and refusing to apologize (Okimoto et al., 2013) involves a kind of pseudo moral repair of a conflict situation, apologizing can be a tool for genuine moral repair (Woodyatt & Wenzel, 2013). Apologies and other signals of remorse (e.g., guilt; shame) improve observers' evaluations of a transgressor's moral character (Stearns & Parrott, 2012). Apologizing also attenuates transgressors' feelings of guilt and shame, and

increases their gratitude, hope, and self-forgiveness (Fisher & Exline, 2010; Hall & Fincham, 2005; Whitvliet et al. 2002; Woodyatt & Wenzel, 2014). However, the extent to which apologizing bolsters transgressors' moral self-image remains unclear. After apologizing, is a transgressor's moral self-image restored back to equilibrium, does it remain somewhat attenuated, or is it perhaps even higher than equilibrium? To test this question, a pre-post design is required. I therefore address this gap in the literature in Study 1 by examining the extent to which transgressors' moral self-image improves after apologizing (vs. not apologizing) compared to baseline and post-offense levels. To the extent that transgressors who apologize experience a boost to their moral self-image, apologizing could influence their subsequent moral behavior in meaningful ways.

1.2 Moral Licensing and Moral Consistency as Consequences of Apologizing

Moral licensing occurs when an individual's prior good deeds allow them the flexibility to engage in immoral, problematic, or less prosocial behavior in the future (Merrit et al., 2010). Moral licensing is relevant in situations where an individual is conflicted between a self-interested choice and a moral choice (Mullen & Monin, 2015). An individual's prior good behavior might free them to act less morally in the future, pursuing self-interested rather than moral behavior without feeling bad or self-punitive.

One psychological mechanism through which moral licensing occurs is through moral credits (see Merritt et al., 2010 for review of mechanisms underlying moral licensing effects). Merritt and colleagues (2010) describe moral credits in terms of a moral bank account, where an individual's prior good deeds earn them credits that they can spend on various subsequent immoral behaviors—even those across different moral domains—that are similar in "moral value." Here,

an individual is tempted between a self-interested choice (e.g., keeping money for self) and a moral choice (e.g., donating money to charity), and their moral credits allow them to choose the former despite recognizing that it is the less moral choice. For instance, Conway and Peetz (2012) found that participants gave less to charity after recalling a past moral behavior.

To the extent that apologies boost moral self-image, this could potentially morally license an individual to act less morally in a different situation. The boost in moral self-image that transgressors might experience following an apology would manifest as moral credits, allowing them to subsequently offend a different victim[1] or withhold prosocial behavior. I investigated these possibilities in Studies 2 and 3.

I anticipate that this moral licensing effect might be most likely to occur under conditions that heighten the perceived morality of the apology. Specifically, I anticipate that apologizing is more likely to earn moral credits for the transgressor when the apology is offered in an ambiguous offense context. Real-world offenses often involve ambiguity around transgressor and victim roles (Thai et al., 2021), and these roles are more fluid than how they are typically studied in experimental settings. A more ambiguous interpersonal offense context might involve both people transgressing to some degree, uncertainty around who should apologize first and who deserves an apology, or a situation where apologizing might not feel completely warranted. Apologizing in these more ambiguous offense contexts might result in greater boosts to transgressors' moral self-image, because apologizing might feel more like a moral choice rather than an obligatory behavior that anyone would perform.

[1] It is also possible that moral licensing via apologies could lead to re-offense of the same victim. However, I will not test this question in the proposed research.

Additionally, I anticipate that apologizing is more likely to earn moral credits for the transgressor when they offer a high-quality apology. High-quality apologies more comprehensively address the psychological needs of the victim (e.g., accepting responsibility; acknowledging harm; offering repair), and contain fewer defensive strategies that serve to protect the transgressor (e.g., denying or minimizing the offense; offering excuses) (see Schumann, 2014 for framework). A higher-quality apology is a more effortful attempt to redress the wrongdoing and reconnect with the victim (Schumann, 2018). A higher-quality apology is also costlier in that it draws more attention to the wrongdoing and the transgressors' involvement in it, and sometimes contains offers for reparations or commitments to change to one's behavior. Because higher-quality apologies are more effortful and costly for the transgressor, they might be seen as a more moral response and therefore should be especially well-suited for repairing a transgressor's moral self-image. Given the potential impact of apology quality, in Studies 1 and 2 I kept apology quality constant by having participants imagine offering a moderate-quality apology. In Study 3, I then allowed apology quality to vary and tested its association with both moral self-image and moral behavior.

However, rather than promote moral licensing, costly prosocial behavior such as an apology might signal strong identification with moral values and thus promote moral consistency (Gneezy et al., 2012). Moral consistency occurs when prior good behavior begets further good behavior. According to self-perception theory (Bem, 1972), people make inferences about their attitudes and values based on their moral behavior. A transgressor's apology—especially a higher-quality apology—might make their moral identity salient and signal their commitment to being a good person. As such, it is possible that apologizing could lead to consistency instead of licensing, enhancing rather than inhibiting transgressors' likelihood of making the moral choice when given

the opportunity. Supporting this possibility, Gneezy and colleagues (2012) examined whether costly prosocial behavior predicts greater subsequent prosocial behavior by manipulating the costliness of charity donations, which included three conditions: 1) deducting money from each participant's compensation to make a donation to a charity on their behalf (i.e., costly), 2) donating money to a charity independent of each participant's compensation (i.e., costless), or 3) giving each participant their compensation without a donation (i.e., control). They found that costly prosocial behavior was associated with feeling moral, which in turn was associated with greater honesty during an economic game.

Although the literature reviewed above supports the plausibility of both moral licensing and moral consistency effects, I expected moral licensing to emerge. Because moral behavior often involves exercising self-control or inhibiting selfish motivations, people can use a prior a good deed to counterbalance their selfish behaviors without negatively impacting their moral self-image (Hofmann et al., 2018). I therefore expected that transgressors would capitalize on having a restored or bolstered moral self-image by feeling free to pursue more self-interested behavior.

However, although I predicted that people will generally be more likely to show licensing than consistency effects, I anticipated the possibility that some people might be especially likely to pursue moral consistency after apologizing. Specifically, people who have a stronger moral identity might be particularly motivated to support this identity with consistent moral behavior (Bem, 1972; Effron & Conway, 2015; Gneezy et al., 2012; Mullen & Monin, 2016). Similarly, I explored the possibility that people who apologize very frequently would feel less morally bolstered by apologizing because to them apologizing simply feels like the default or appropriate thing to do. If so, these high frequency apologizers might be less likely to show moral licensing

effects. To test these possibilities, I explored interactions between apologizing condition and trait levels of moral identity and apology proclivity in Studies 2 and 3.

1.3 Moral Cleansing as a Consequence of Not Apologizing

My prediction that apologizing can lead to moral repair also potentially carries implications for behavioral consequences of *not* apologizing. How does refraining from moral repair via an apology after committing an interpersonal offense impact a transgressor's subsequent moral behavior? Moral cleansing occurs when people engage in moral behavior as a way to restore their damaged moral self-image following a bad deed (Merritt et al., 2010). Transgressors who do not morally cleanse via an apology might consequently compensate by engaging in other forms of moral behavior. This moral cleansing effect might be especially likely to occur in situations where the transgressor recognizes their moral responsibility for an offense but still does not apologize (e.g., because they think it will be uncomfortable to do so, Leunissen et al., 2014; Schumann, 2018; because they don't have an opportunity to apologize). I investigated this possibility in Studies 2 and 3 by examining the effects of not apologizing on subsequent moral behavior. Moreover, because not apologizing might be especially morally uncomfortable for people with either a strong moral identity or a strong tendency to apologize, I explored interactions between apologizing condition and both trait levels of moral identity and trait levels of tendency to apologize.

1.4 Research Overview

In three pre-registered studies (https://osf.io/69thk/?view_only=8e9aa8aadd27475d9e5b3fe5a006a148), I examined how apologizing (vs. not apologizing) for an interpersonal offense affects transgressors' moral self-image and subsequent moral behavior. In Study 1, I used a repeated measures design to examine the extent to which apologies repair participants' moral self-image. After committing an offense, do apologies slightly increase, fully restore, or even enhance transgressors' moral self-image above baseline levels? Participants completed a baseline measure of moral self-image and then imagined an interpersonal interaction where they commit an offense against a roommate. Moral self-image was assessed again to test whether transgressors experience a loss of perceived morality after committing an offense. Next, participants imagined either apologizing or not apologizing, then moral self-image was assessed a third time to assess (a) the extent to which apologizing repairs their moral self-image relative to both post-offense/pre-apology levels and the non-apology condition, and (b) the extent to which not apologizing leaves people in a morally depleted state.

In Studies 2 and 3, I used complementary methods to examine the implications of these changes to moral self-image on transgressors' subsequent moral behavior. Specifically, I examined whether apologies can morally license transgressors to behave less morally, and whether or not apologizing motivates transgressors to morally cleanse by engaging in moral actions. In Study 2, I tested these effects using a hypothetical offense to standardize the nature and severity of the offense, as well as the content of the apology. In Study 3, I used a recalled offense to increase ecological validity, and instructed participants to write their own apology to the victim. The present work extends the literature by examining (a) how apologies impact transgressors' moral self-image, which has not been clearly documented in the prior literature, and (b) the psychology that

persists and spills over after an apology to impact transgressors' subsequent moral behavior in the form of moral licensing and moral cleansing—both of which are novel outcome measures in the context of apology behavior. This work will therefore offer insights into whether there are conditions under which apologies might backfire, freeing transgressors up to withdraw from subsequent opportunities to engage in moral behavior.

2.0 Study 1

In Study 1, I examined how apologizing vs. not apologizing repairs moral self-image following a hypothetical offense. Whereas moral identity tends to be a stable characteristic that refers to an individual's "self-conception organized around a set of moral traits" (Aquino & Reed, 2001, p. 1424), moral self-image refers to an individual's "malleable self-concept" (Jordan et al., 2015, p. 3) that can shift based on cues from other people and from perceptions of one's moral actions. As such, an individual might negatively evaluate their moral self-image following an offense and positively evaluate their moral self-image following an apology. Supporting this possibility, Jordan and colleagues (2015) demonstrated that recalling a moral behavior increases one's moral self-image whereas recalling an immoral behavior decreases one's moral self-image. In Study 1, participants read a hypothetical offense and then were randomly assigned to one of two conditions where they either imagined apologizing or not apologizing. Moral self-image was measured at baseline, post-offense, and post-response (apology vs. no apology). I hypothesized the following:

Hypothesis 1: Within-Subject Change Post-Offense. Participants' moral self-image will be significantly lower post-offense (t2) compared to baseline (t1).

Hypothesis 2: Within-Subject Change Post-Apology. Participants' moral self-image will be significantly higher post-apology (t3) compared to post-offense (t2).

Hypothesis 3: Post-Apology vs. Post-No Apology. Moral self-image ratings will be significantly greater post-apology (t3) compared to post non-apology (t3).

In addition to testing these primary hypotheses, I explored *the extent* to which apologizing vs. not apologizing causes moral repair or moral damage by testing the within-subject changes between each time point within the apology and no apology conditions.

2.1 Method

2.1.1 Participants

A power analysis was conducted using the R software package Superpower (Lakens & Caldwell, 2021). The analysis suggested that a sample size of at least 35 participants per cell (N = 210) would be sufficiently powered ($\alpha = .05$) to detect the hypothesized main effect for condition (estimated $\eta^2 = .15$), main effect for time (estimated $\eta^2 = .15$), and interaction between time and condition (estimated $\eta^2 = .41$). Additionally, to examine the simple effects, the analysis suggested a sample size of 35 participants per cell to detect a medium effect size ($d = .50$, $\alpha = .05$, 80% power). Because the offense scenario involved a roommate, I first prescreened people on Prolific to recruit participants who were currently living with one or more roommate(s). Five hundred fifty-three participants completed the prescreen question and were compensated 25 cents for their time. Two hundred sixty-three participants indicated that they currently live with a roommate and were invited to complete the main study. Two hundred forty-three participants completed the survey and were compensated $1.25 for their time. All participants met the preregistered data inclusion criteria. The final sample of 243 participants consisted of 124 women, 117 men, 1 non-binary/third gender, and 1 unspecified (M_{age} = 24.96, SD = .3.49; $Median_{education}$ = 4-year degree). The

distribution of race/ethnicity was 76.54% White, 13.17% Asian, 10.29% Latinx, 5.76% Black/African American, mixed race/ethnicity 6.58%, and Middle Eastern/North African, .82%.

2.1.2 Materials and Procedure

2.1.2.1 Moral Self-Image

Participants first completed an adapted version of the moral self-image scale (Aquino & Reed, 2001; Jordan et al., 2015), which is a 9-item self-report questionnaire assessing an individual's "self-concept related to the traits of a prototypically moral person" (e.g., caring, generous, and honest) (Jordan et al., 2015, p. 3). I modified the instructions such that participants rated how they feel "right now." I also substituted the "hardworking" item and instead used the term "moral" because of its relevance to the context of the hypothetical offense (see Appendix C). The items were answered on a modified 7-point scale (e.g., 1 = *Not at all caring*, 7 = *Extremely caring*) and were averaged at each time point to create an index of moral self-image ($\alpha = .90$).

2.1.2.2 Conflict Scenario

Next, participants were asked to imagine themselves in a conflict situation with their roommate (see Appendix B for full scenario). To promote a sense of ambiguity, the participant and roommate both act in ways that might offend each other. Specifically, participants imagined breaking their promise to their roommate to stick to a cleaning schedule, but also imagined that their roommate had people over late on a weeknight despite knowing the participant needed to get up early for work.

2.1.2.3 Moral Self-Image

Participants then completed the same moral self-image scale, as previously described, a second time. Participants were asked to complete the same questions again based on the scenario they just imagined ($\alpha = .93$).

2.1.2.4 Response to Transgression

Next, participants were randomly assigned to either the apology condition or no apology condition. Participants in the apology condition were asked to imagine themselves texting their roommate a moderate-quality apology—where they express remorse, accept responsibility, and promise to behave better in the future—whereas participants in the no apology condition were asked to imagine themselves considering offering their roommate an apology but then deciding not to (see Appendix C).

2.1.2.5 Moral Self-Image

Participants then completed the same moral self-image scale a third time, based on the continuation of the story they had just imagined ($\alpha = .90$) (see Appendix D, and see Appendix E for additional measures).

2.1.2.6 Demographics

Lastly, participants were asked to indicate the gender of the roommate they imagined in the conflict scenario, completed a demographics questionnaire, and were debriefed via an online feedback letter.

2.2 Results

I conducted a within-between ANOVA, with time of measurement as the within-subjects factor and apology condition as the between-subjects factor, to test hypotheses 1-3. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $W(2) = .994$, $p = .533$. As such, the omnibus statistics reported below are based on sphericity assumed. I applied Bonferroni corrections for multiple comparisons that involved more than two comparisons.

I first tested *hypothesis 1* by examining changes in moral self-image over time. There was a significant main effect of time, $F(2, 482) = 49.61$, $p < .001$, $\eta^2 = .171$ (see Table 1). Consistent with *hypothesis 1*, participants' moral self-image was significantly lower post-offense ($M = 4.18$, $SD = 1.21$) compared to baseline ($M = 4.83$, $SD = 1.06$), $p < .001$, Hedges's G_{av}[2] = .57, 95% CI [.44, .70]. In addition, there was no statistically significant difference in baseline levels ($p = .525$) and post-offense levels ($p = .697$) of moral self-image between the apology and no apology condition, showing that the random assignment procedure was effective, and that the offense impacted participants' moral self-image similarly in both conditions (see Figure 1). That is, changes in moral self-image from baseline to post-offense were virtually identical for participants in the apology and no apology condition.

Next, I tested *hypotheses 2* and *3* by examining the interaction between apology condition and time on moral self-image ratings. There was a significant interaction between time and apology condition, $F(2, 482) = 79.11$, $p < .001$ (see Table 1). Consistent with *hypothesis 2*, within the

[2] Hedges's G_{av} is generally recommended when calculating effect sizes for correlated samples, and it is a comparable estimate to effect sizes for between-subjects designs (Lakens, 2013).

apology condition, participants' moral self-image was significantly higher post-apology (M = 5.30, SD = 1.01) compared to post-offense (M = 4.15, SD = 1.18), $p < .001$, Hedges's G_{av} = 1.04, 95% CI [.83, 1.26], suggesting that apologizing had a reparative effect on moral self-image (see Figure 1). Consistent with *hypothesis 3*, comparing differences between apology conditions at time 3, participants who imagined apologizing report higher moral self-image (M = 5.30, SD = 1.01) than participants who imagined not apologizing (M = 3.92, SD = 1.25), $p < .001$, d = 1.21, 95% CI [.94, 1.49], suggesting that apologizing bolstered moral self-image relative to withholding an apology (see Figure 1).

Moreover, apologizing not only restored moral self-image but also bolstered moral self-image above baseline levels. Participants in the apology condition reported significantly greater levels of moral self-image post-apology (M = 5.30, SD = 1.01) compared to baseline levels (M = 4.48, SD = 1.00), $p <.001$, Hedges's G_{av} = .62, 95% CI [.48, 1.00]. In contrast, not apologizing lowered moral self-image below baseline levels. Participants in the no apology condition reported significantly lower levels of moral self-image after not apologizing (M = 3.92, SD = 1.25) compared to baseline (M = 4.87, SD = 1.13), $p < .001$, Hedges's G_{av} = .80, 95% CI [.46, .98], as well as compared to post-offense (M = 4.21, SD = 1.24), p = .005, Hedges's G_{av} = .30, 95% CI [.14, .64].

2.3 Discussion

In Study 1, I found that the hypothetical offense attenuated transgressors' moral self-image but apologizing had restorative effects. Indeed, apologizing not only restored moral self-image to equilibrium but even bolstered moral self-image above baseline levels. In terms of how this

bolstering effect might impact subsequent moral behavior, these findings are consistent with theorizing on moral licensing such that good behavior bolsters state levels of moral self-image, which in turn leads to less moral subsequent behavior (Conway & Peetz, 2012; Khan & Dhar, 2006; Merritt et al., 2010). I will examine this possibility in Study 2. Further, the results from Study 1 suggest that not engaging in moral repair in the aftermath of an offense can be morally depleting. In Study 2, I will test whether this morally depleted state results in moral cleansing, where people compensate for their prior immoral behavior by subsequently engaging in more moral behavior (Conway & Peetz, 2012; Mullen & Monin, 2016; Sachdeva et al., 2009; Zhong & Liljenquist, 2006).

3.0 Study 2

In Study 2, I examined whether apologizing morally licenses transgressors to offend another person, and whether transgressors who do not morally cleanse via an apology are less likely to offend another person. The same hypothetical offense in Study 1 was used in Study 2. However, rather than use a pre-post design, I included a no-offense control condition to examine whether licensing or cleansing is occurring and to examine the magnitude of each effect (Merritt et al., 2010).

Following the initial hypothetical offense and apology scenario, participants read a second hypothetical workplace scenario. This scenario involved being tempted to act in a more self-interested vs. prosocial way, where participants could take credit for their co-worker's ideas to help them earn a promotion. To the extent that apologizing earns them moral credits, I hypothesized that participants in the apology condition (vs. no apology and control conditions) would be more likely to take credit for their coworker's ideas (i.e., moral licensing). Moreover, to the extent that withholding an apology results in moral depletion, I hypothesized that participants in the no apology condition (vs. apology and control conditions) would be less likely to take credit for their coworker's ideas (i.e., moral cleansing). Thus, I hypothesized the following:

Hypothesis 1: Moral Licensing. Participants in the apology condition will be more likely to take credit for their coworker's ideas compared to participants in the no apology and control conditions.

Hypothesis 2: Moral Cleansing. Participants in the no apology condition will be less likely to take credit for their coworker's ideas compared to participants in the apology and control conditions.

3.1 Method

3.1.1 Participants

A power analysis (Lakens & Caldwell, 2021) suggested that a sample size of at least 65 participants per cell (N = 195) would be sufficiently powered to detect a main effect based on the hypothesized overall effect size and mean differences (α = .05, estimated η^2 = .15). Additionally, to examine the simple effects, the analysis suggested 65 participants per cell to detect medium sized effects (d = .50, α = .05, 80% power).

Nine hundred fifty-five participants were first prescreened to be currently living with one or more roommates(s) and were compensated 25 cents. From this prescreen, two hundred fifty-seven participants were invited to participate in the main study, 216 of which completed the survey and were compensated $3.00 for their time. Five participants failed the attention check and were excluded from all analyses. The final sample of 211 participants consisted of 104 women, 104 men, 2 non-binary/third gender, and 1 unspecified (M_{age} = 26.31, SD = 4.99; $Median_{education}$ = 4-year degree). The distribution of race/ethnicity was 55.92% White, 19.43% Asian, 8.53% Latinx, 8.53% mixed race/ethnicity, 7.58% Black/African American, .95% Native American, and .95% other race/ethnicity.

3.1.2 Materials and Procedure

3.1.2.1 Experimental Conditions

Participants were first randomly assigned to one of three conditions: apology, no apology, or a no offense control condition. Participants in the apology and no apology conditions read the

same scenario from Study 1, whereas participants in the control condition read a scenario that did not involve offensive behaviors—efforts were taken to keep the details as similar as possible to the apology vs. no apology conditions while removing the offensive behaviors (see Appendix F for a full description of the scenarios).

3.1.2.2 Moral Self-Image

Participants then completed the same moral self-image scale ($\alpha = .95$) as in Study 1 (see Appendix D).

3.1.2.3 Hypothetical Workplace Scenario

Next, participants in each condition imagined themselves in a workplace scenario where they could take credit for their co-worker's ideas (see Appendix G for full description). To measure likelihood of moral behavior, participants were asked how likely they would be to give their co-worker credit for their contributions to the task (1 = *Not at all likely*, 7 = *Extremely likely*), where higher scores indicated a more moral decision. To assess their perceptions of the morality of this decision, participants were also asked to rate how immoral it would be to not speak up, thereby accepting credit for their coworker's contributions to the task (1 = *Not at all immoral*, 7 = *Extremely immoral*).

3.1.2.4 Offense Perceptions

To better understand how participants perceived the original offense situation with their roommate, they completed items assessing self-negativity ("to what extent did you behave negatively toward Taylor?"), other-negativity ("to what extent did Taylor behave negatively toward you?"), self-apology deservingness ("to what extent did you deserve an apology from

Taylor?") and other-apology deservingness ("To what extent did Taylor deserve an apology from you?"), which were answered on a 7-point scale (e.g., 1 = *Not at all/To a very small extent,* 7 = *Extremely/To a very large extent*). Additionally, participants in the apology and no apology condition completed two items assessing the severity of both their actions and their roommate's actions (e.g., "How severe is what you did to Taylor?"), which were answered on a 7-point scale (1 = *Not at all severe*, 7 = *Very severe*) (see Appendix H).

3.1.2.5 Manipulation Check

Participants were then asked to what extent they were apologetic in the scenario with Taylor (1 = *Not at all apologetic*, 7 = *Extremely apologetic*) and accepted responsibility for their actions (1 = *To a very small extent*, 7 = *To a very large extent*) (see Appendix H).

3.1.2.6 Exploratory Moderators

In random order, participants completed the proclivity to apologize measure (PAM) (Howell et al., 2011) and the moral identity questionnaire (Black & Reynolds, 2016). The PAM is an 8-item self-report questionnaire (e.g., "I don't apologize very often because I don't like to admit that I'm wrong") assessing a person's general willingness to apologize. The items were answered on a 7-point scale (1 = *Strong disagreement*, 7 = *Strong agreement*), and composite scores were created by averaging the items (α = .91) (see Appendix I).

The moral identity questionnaire is a 20-item self-report questionnaire (e.g., "I try hard to act honestly in most things I do.") that consists of two subscales—moral integrity and moral self—assessing "the importance people give both to their moral principles and to acting accordingly, independently of religious or political views" (p 121). The items were answered on a 7-point scale

(1 = *Strongly disagree*, 7 = *Strongly agree*), and composite scores were created by averaging all the items (α = .90) (see Appendix I).

3.1.2.7 Demographics

Lastly, participants completed a demographics questionnaire and were then debriefed via an online feedback letter.

3.2 Results

3.2.1 Preliminary Analyses

I first examined differences between conditions on the manipulation check questions and in perceptions of the offense (see Table 2 for correlations between all dependent measures and Table 3 for means by condition). As expected, participants in the apology condition rated themselves as more apologetic ($p < .001$) and accepted more responsibility for their actions ($p < .001$) than participants in the no apology condition. Participants in the apology and no apology condition rated both their own actions and Taylor's actions as more negative ($ps \leq .001$) than participants in the control condition, and felt that both themselves and Taylor deserved an apology to a greater extent than participants in the control condition ($ps \leq .001$). However, there were no statistically significant differences between the apology and no apology condition on self- and other-negativity ratings ($ps \geq .077$) or self- and other-apology deservingness ratings ($ps = 1.00$), and there were no differences between the apology and no apology condition in the severity of

their own actions and of Taylor's actions (*p*s ≥ .051). Thus, apologizing vs. not apologizing did not seem to impact participants' perceptions of the offense.

3.2.2 Primary Analyses

3.2.2.1 Differences in Moral Self-Image between Conditions

Next, I used a one-way ANOVA to examine differences in moral self-image ratings between the three conditions. I applied Bonferroni corrections to analyze pairwise comparisons. Results showed a significant main effect of condition on moral self-image ratings, $F(2, 208) = 20.86$, $p < .001$, $\eta^2 = .167$ (see Table 3 and Figure 2). Replicating Study 1, participants in the apology condition ($M = 4.97$, $SD = 1.08$) reported greater levels of moral self-image compared to participants in the no apology condition ($M = 4.02$, $SD = 1.26$), $p < .001$, $d = .81$, 95% CI [.47, 1.16]. Additionally, participants in the control condition ($M = 5.20$, $SD = 1.08$) reported greater levels of moral self-image compared to participants in the no apology condition, $p < .001$, $d = 1.01$, 95% CI [.65, 1.36]. However, there was no statistically significant difference in moral self-image ratings between the apology and control condition, $p = .705$, $d = .21$, 95% CI [-.12, .54], suggesting that in this study, apologizing restored moral self-image to baseline levels but did not bolster beyond that.

3.2.2.2 Moral Licensing and Moral Cleansing

To test *hypotheses 1* (i.e., moral licensing) and *2* (i.e., moral cleansing), I used a one-way ANOVA to examine differences in moral licensing between the three conditions. I applied Bonferroni corrections to analyze pairwise comparisons. Contrary to my hypotheses, there was no main effect of condition on taking credit for a coworker's idea, $F(2, 208) = 1.64$, $p = .197$, $\eta^2 =$

.016 (see Table 3 and Figure 3). As such, there were no differences in moral licensing or moral cleansing among the conditions.[3] However, based on the directionality of the means, results might suggest that participants trended toward moral cleansing after withholding an apology and moral consistency after apologizing (see Figure 3). Additionally, there was no statistically significant difference between the conditions in ratings of how immoral it would be to take credit for the coworker's contributions, $F(2, 207) = 2.54$, $p = .081$).

3.2.2.3 Mediation via Moral Self-Image

Next, I examined the indirect effects of apology condition on moral behavior through moral self-image. I used PROCESS, model 4 with 10,000 bias-corrected bootstrap samples (Hayes, 2017) and included orthogonal contrasts (i.e., Helmert coding)apology vs. no apology and control (X1) and no apology vs. control (X2) as predictors[4], and moral self-image as the mediator (M). The indirect effects were not significant (apology vs. no apology and control indirect effect: -.03, 95% CI [-.12, .04]; no apology vs. control indirect effect: -.08, 95% CI [-.30, .14]). This was not surprising, as moral self-image did not significantly correlate with moral behavior, $r(209) = .05$, $p = .512$.

[3] The distribution for the moral licensing measure was negatively skewed and leptokurtic and outliers were present, particularly in the control and no apology conditions (see Figure 2). I re-analyzed the data after removing the outliers (i.e., five cases had z-scores less than -3.29) and transforming the data (i.e., reflect and square root). However, these changes did not impact the interpretation of the omnibus ANOVA ($p = .229$) or pairwise comparisons ($ps \geq .375$).

[4] The following orthogonal contrasts using effect size coding were used: Apology vs. no apology and control (-.66, .33, .33), no apology vs. control (0, .5, -.5).

3.2.3 Exploratory Analyses

3.2.3.1 Interactions with Moral Identity, PAM, Negativity, and Apology Deservingness

Finally, I explored whether moral identity, PAM, and perceptions of the offense moderated the effects of condition on the decision to take credit for their coworker's idea. I used orthogonal contrasts to compare apology vs. no apology and control and no apology vs. control using effect size coding[5]. I ran separate multiple regressions that included moral identity, PAM, self-negativity, other-negativity, self-apology deservingness and other-apology deservingness as exploratory moderators (e.g., the model examining the moderating role of moral identity included the following predictors: D2: apology vs. no apology and control; D2: no apology vs. control; mean-centered moral identity, D1*moral identity; D2*moral identity).

As can be seen in Table 4, participants with higher levels of moral identity or a greater proclivity to apologize were more likely to give their coworker credit. However, neither of these trait measures moderated the effect of the apology condition on moral licensing ($ps \geq .160$) or moral cleansing ($ps \geq .082$). Regarding other-negativity ratings and other-apology deservingness, neither measure predicted whether participants took credit for their coworker's idea ($ps \geq .145$) and neither moderated the effect of the apology condition on moral licensing ($ps \geq .208$) or moral cleansing ($ps \geq .169$) (see Table 4 for full regression model).

However, there was an interaction between self-negativity ratings and the no apology vs. control contrast ($b = .39$, $p = .027$). For those who rated their own behavior more negatively (i.e.,

[5] The following orthogonal contrasts using effect size coding were used: Apology vs. no apology and control (-.66, .33, .33), no apology vs. control (0, .5, -.5).

+1 *SD* above the mean), participants in the no apology condition were more likely to give their coworker credit than participants in the control condition (b = 1.04 , p = .021) (see Figure 4 and Table 4). These results suggest that those who viewed their offense as highly negative and did not apologize were more likely to morally cleanse than those in the control condition. However, at low levels of self-negativity, there was no difference in moral behavior between the no apology and control condition (b = -.24, p = .523). There was no interaction between the apology vs. no apology contrast and self-negativity ratings (b = .08, p = .574). There was a main effect for the apology vs. no apology and control contrast (b = -.58, p .010), such that participants in the apology condition were more likely to give credit than participants in the no apology and control condition at average levels of self-negativity. The average slope for self-negativity across each condition was not statistically significant (b = -.08, p = .238).

3.3 Discussion

In Study 2, I did not find evidence that apologizing impacts subsequent moral behavior in the form of moral licensing and moral cleansing. Although apologizing had a positive effect on moral self-image ratings, replicating findings from Study 1, this did not seem to spill over and license participants to make the less moral choice of accepting credit for a coworker's ideas. If anything, there was some evidence of a moral consistency effect after apologizing. Additionally, withholding an apology had a negative effect on moral self-image ratings, replicating findings from Study 1, but this did not seem to spill over and motivate moral cleansing, except under high levels of perceived negativity of the offense. Thus, it could be that moral cleansing is most likely

to occur when participants believe they have a lot to apologize for but then do not offer that apology.

What might account for the lack of moral licensing effect? It seems possible that the subsequent behavior of taking ownership of a coworker's idea for personal gain was too immoral or unethical. Participants tended to highly endorse giving their coworker credit and perceived taking credit for their coworker's ideas as very immoral. As such, the moral credits earned via an apology might have been incommensurate with the cost of the subsequent immoral behavior. Apologizing might not have been enough to license participants to offend their coworker. In Study 3, I conducted a more highly powered and ecologically valid study, using a different operationalization of moral licensing (i.e., donations) to address some of the methodological limitations of Study 2.

4.0 Study 3

In Study 3, I used complementary methods to further examine whether apologizing vs. not apologizing impacts subsequent moral behavior. Participants were asked to recall an unresolved conflict and then were either asked to write an apology letter or took no further action and moved on to subsequent measures. A no-offense condition was also included where participants were not asked to recall a conflict. Toward the end of the study, participants were given bonus money for their participation and decided if they would like to donate none, all, or some portion of their bonus to a non-profit organization. The donation amount has been used as a prosocial outcome in prior work on moral licensing (e.g., see Conway & Peetz, 2012). To the extent that apologizing bolsters moral self-image and leads to moral licensing, I expected participants in the apology condition to donate less of their bonus compared to participants in the no apology and control conditions. Thus, I hypothesized the following:

Hypothesis 1: Moral Licensing. Participants in the apology condition will donate significantly less money than participants in the no apology and control conditions.

Hypothesis 2: Moral Cleansing. Participants in the no apology condition will donate significantly more money than participants in the apology and control conditions.

4.1 Method

4.1.1 Participants

A power analysis (Lakens & Caldwell, 2021) suggested that a sample size of 210 participants per cell (N = 630) would be sufficiently powered to detect a main effect based on the size of the overall effect of the apology condition on moral behavior found in Study 2 (80%, α = .05, estimated η^2 = .02). Additionally, to examine the simple effects, the analysis suggested 210 participants per cell to detect small sized effects (d = .31, α = .05, 88% power).

Six-hundred thirty participants were sampled and compensated $2.50 for their participation (not including their potential bonus payment of up to $1.00; see below for a full description of the procedures). Two participants failed the attention check and were excluded from all analyses, and 35 participants were excluded from all analyses because they did not follow the apology letter instructions. The final sample of 593 participants consisted of 284 women, 291 men, 16 non-binary/third gender, and 2 unspecified (M_{age} = 35.47, SD = 12.50; $Median_{education}$ = 4-year degree). The distribution of race/ethnicity was 65.90% White, 9.11% Asian, 8.43% Black/African American, 6.75% Latinx, 8.94% mixed race/ethnicity, .67% Middle Eastern or North African, .34% Native American, .34% other race/ethnicity, and .17% unspecified.

4.1.2 Materials and Procedure

4.1.2.1 Experimental Conditions

Participants were randomly assigned to one of three conditions: apology, no apology, or control condition. In the apology and no apology conditions, participants were asked to think of

an offense that was currently unresolved (see Appendix J for recall prompt and instructions for each condition). Participants in the apology condition were asked to write a short description of the offense and then were asked to write a letter to the person they offended, with a special request that they use this opportunity to apologize (adapted from Schumann & Walton, 2021). Participants in the no apology condition were asked to provide a short description of the offense. Participants in the control condition were asked to recall the last interaction they had with someone and to provide a short description of the interaction.

To explore the role of apology quality in moral licensing, each participant in the apology condition rated the quality of their apology letter. Specifically, participants self-rated their level of agreement with eight items that assessed comprehensiveness (e.g., "I accepted responsibility for what happened") and four items that assessed defensiveness (e.g., "I justified my behavior"), which were answered on a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*) (see Appendix K).

4.1.2.2 Moral Self-Image

Participants then completed the same moral self-image scale as in Study 1 ($\alpha = .94$) (see Appendix D).

4.1.2.3 Donations

Next, participants were given a $1.00 bonus for their participation. The instructions stated that participants would be paid $2.50 for their participation but would be given a bonus as a result of additional funding received for the study. Participants could keep this money for themselves or could donate none, all, or some portion of the bonus money to GiveWell's Maximum Impact Fund, which is a non-profit organization that donates to charities that save or improve lives the most per

dollar. The instructions stated that this decision was completely up to them. Participants were then asked how much money they would like to donate (*0 cents – $1.00*) (see Appendix L). Participants actually received the amount of bonus they chose to keep.

4.1.2.4 Closeness and Negativity Ratings

Participants in all three conditions were asked how close they were to the person at the time of the interaction, which was answered on a 7-point scale (1 = *Not at all close*, 7 = *Extremely close*). Participants in all three conditions also completed three questions that assessed how negative the interaction was (e.g., "How offensive was the interaction?"), plus three positive filler items, each of which was answered on a 7-point scale (e.g., 1 = *Not at all offensive*, 7 = *Extremely offensive*). The items assessing negativity were combined to create a composite score ($\alpha = .90$) (see Appendix J).

4.1.2.5 Exploratory Moderators

In random order, participants then completed the same PAM ($\alpha = .91$) and moral identity questionnaire ($\alpha = .91$) as in Study 2 (see Appendix I).

4.1.2.6 Demographics

Lastly, participants completed a demographics questionnaire and an open-ended suspicion check regarding the purpose of the study, and were then be debriefed via an online feedback letter.

4.2 Results

4.2.1 Preliminary Analyses

Overall, participants rated their recalled offenses as moderately negative (M = 3.72, SD = 2.03). Negativity ratings in the apology (M = 4.88, SD = 1.35) and no apology condition (M = 4.97, SD = 1.30) were significantly higher than the control condition (M = 1.64, SD = 1.18), $F(2, 590) = 457.27$, $p < .001$, $\eta^2 = .608$ (see Table 5 for correlations between all dependent measures and Table 6 for means by condition). However, there was no significant difference in negativity ratings between the apology and no apology condition (p = 1.00).

Participants reported feeling moderately close to the person at the time of the interaction (M = 5.08, SD = 1.76), but closeness ratings did not statistically differ among the conditions, $F(2, 590) = .53$, $p = .588$, $\eta^2 = .002$. The offenses recalled involved criticism (24%), disagreements (13%), failed obligation (11%), betrayal (4%), deception (3%), and possession (3%), inconvenience (1%), ambiguous or unclear offense (.3%), and physical offense (.2%).

4.2.2 Primary Analyses

4.2.2.1 Differences in Moral Self-Image between Conditions

I used ANOVA to examine differences in moral self-image ratings between the three conditions. I applied Bonferroni corrections for multiple comparisons. Results showed a significant main effect of condition on moral self-image ratings, $F(2, 590) = 11.30$, $p < .001$, $\eta^2 = .037$ (see Table 6 and Figure 5). However, these findings were not in the expected direction based on findings from Studies 1 and 2. Participants in both the apology (M = 4.84, SD = 1.12), p = .004,

and no apology condition (M = 4.72, SD = 1.26), $p < .001$) reported significantly lower levels of moral self-image than participants in the control condition (M = 5.22, SD = .99). There was no difference in moral self-image ratings between the apology and no apology condition, $p = .870$.

Looking within the apology condition only, apology comprehensiveness [$r(159) = .12$, $p = .147$)] and defensiveness [$r(159) = -.09$, $p = .235$)] were not significantly associated with moral self-image. Although statistical power was reduced by examining these relationships only within the apology condition, this suggests that higher quality apologies were not associated with higher levels of moral self-image in this study.

4.2.2.2 Moral Licensing and Moral Cleansing

Next, I used ANOVA to examine differences between conditions in donations. The main effect of apology condition on donations was not significant, $F(2, 590) = .93$, $p = .395$, $\eta^2 = .003$ (see Table 6 and Figure 6). Again, examining only the apology condition, apology comprehensiveness [$r(159) = .12$, $p = .147$)] and defensiveness [$r(159) = -.09$, $p = .235$)] were not significantly associated with donations (see Table 5).

4.2.2.3 Mediation via Moral Self-Image

I then examined the indirect effects of apology condition on donations through moral self-image. I used PROCESS, model 4 with 10,000 bias-corrected bootstrap samples (Hayes, 2017) and included orthogonal contrasts (i.e., Helmert coding) apology vs. no apology and control (X1) and no apology vs. control (X2) as predictors[6], and moral self-image as the mediator (M). Moral

[6] The following orthogonal contrasts using effect size coding were used: Apology vs. no apology and control (-.66, .33, .33), no apology vs. control (0, .5, -.5).

self-image was significantly correlated with donations, $r(591) = .11$, $p = .006$. However, the indirect effects were not significant (apology vs. no apology and control indirect effect: 0.00, 95% CI [0.00, .01]); no apology vs. control indirect effect: -.02, 95% CI [-.03, 0.00].

4.2.3 Exploratory Analyses

4.2.3.1 Interactions with Moral Identity, PAM, Negativity, and Closeness

Finally, I ran separate multiple regressions with mean-centered moral identity and PAM as predictors to examine whether each variable moderated the effect of the apology condition on donations. Regarding the main effects, moral identity ($b = .06$, $p = .006$) and PAM ($b = .04$, $p < .001$)predicted donations (see Table 7 for full regression models). Neither moral identity nor PAM moderated the effect of the apology condition on donations. I also explored whether mean-centered negativity and mean-centered closeness moderated the effect of apologizing on donations. However, negativity and closeness were not significant moderators (see Table 7).

4.3 Discussion

In Study 3, I did not find evidence that apologizing licensed participants to donate less bonus money relative to either withholding an apology or no offense controls. Additionally, I did not find evidence that withholding an apology influenced participants to morally cleanse by donating more bonus money relative to either apologizing or no offense controls. Thus, apologizing vs. not apologizing did not seem to impact subsequent moral behavior, contrary to my hypotheses, even under conditions of high negativity (as was seen in Study 2).

Study 3 methods were intended to address the limitations of the hypothetical nature of the offense, response to the transgression, and moral licensing scenarios used in Studies 1 and 2. However, the apology letters seemed to backfire because the recalled offenses tended to have a sustained, adverse effect on participants' moral self-image across apology conditions. Although participants decided what to write in their apology letters, apologizing was not a spontaneous behavior and might not have been experienced as a genuine or sincere attempt to restore the situation. Indeed, prior work suggests that an externally motivated apology could reduce its licensing power because voluntary behaviors are more likely to bolster moral self-image than involuntary behaviors (e.g., see Mullen & Monin, 2016). As a result, participants in the apology and no apology condition reported lower levels of moral self-image compared to participants in the control condition.

There also seemed to be limitations to using donations as a measure of moral licensing. Fifty-two percent of participants chose to donate none of their bonus money. It is possible that participants viewed their decision to donate as a financial decision as opposed to a moral decision. Supporting this interpretation, donations were only very modestly associated with moral self-image and moral identity, suggesting that factors other than participants' moral self-image were governing their donation decision.

5.0 General Discussion

While conflict may be inevitable across all types of relationships, apologizing can have a powerful effect on restoring relationship harmony and addressing the psychological needs of both victims and transgressors (Fehr et al., 2010; Fisher & Exline, 2010; Forster et al., 2021; Hall & Fincham, 2005; Schnabel & Nadler, 2008; Schumann & Dragotta, 2021; Witvliet et al., 2001). Yet the extent to which apologizing repairs transgressors' moral self-image is not well documented, and little is known about how apologizing impacts transgressors' subsequent moral behavior. If apologizing repairs (and possibly bolsters) transgressors' moral self-image whereas withholding an apology precludes repair to moral self-image, these conflict responses could impact transgressors' subsequent moral behavior. Although I anticipated the possibility of morally consistency effects, I predicted that an apology would license transgressors to act less morally or withhold prosocial behavior, whereas withholding an apology would lead transgressors to morally cleanse by behaving more morally toward others.

Across Studies 1 and 2, apologizing had restorative effects on participants' moral self-image whereas withholding an apology left participants in a morally depleted state. However, Studies 2 and 3 did not demonstrate that apologizing vs. not apologizing impacted subsequent moral behavior in the form of moral licensing and moral cleansing.

The highlight of Studies 1 and 2 was the consistent effect that apologizing vs. not apologizing had on participants' moral self-image. Prior work has shown that apologies attenuate negative emotions within transgressors following an interpersonal offense, such as guilt and shame (Fisher & Exline, 2010; Hall & Fincham, 2005). The present findings contribute to the literature by focusing on moral self-image—showing that apologizing repaired the negative effect that the

offense had on transgressors' moral self-image, whereas not apologizing resulted in participants feeling morally depleted. Moral self-image can be damaged following an interpersonal offense (Schnabel & Nadler, 2008), and transgressors might even engage in various forms of self-punishment to communicate remorse and restore the moral values that were violated because of their wrongdoing (Hechler et al., 2022). As such, it is important to document the extent to which a more constructive and restorative approach—namely apologizing—can repair a person's moral self-image in the aftermath of a wrongdoing. The present findings suggest that apologizing can restore—and possibly even bolster—transgressors' moral self-image, thus restoring an important element of their identity (Stromingher & Nichols, 2014) that was impaired by the act of offending someone (Schnabel & Nadler, 2008).

Despite the impact that apologizing vs. not apologizing had on participants' moral self-image, this did not seem to impact their subsequent moral behavior. Moral self-regulation theory posits that people can earn moral credits by engaging in good behavior that could then license them to engage in less moral subsequent behavior, thus balancing out the negative effect that an immoral act can have on their moral self-image (Mullen & Monin, 2016). In contrast, moral cleansing occurs when people balance out their past immoral behavior by engaging in good subsequent behavior. However, the current studies did not show strong evidence of moral licensing or moral cleansing. What might account for these null effects?

As previously mentioned, it is possible that limitations with the paradigms used played a role in producing null effects. However, another possibility is that apologizing might not license transgressors to offend a different person. The boost in moral self-image after apologizing might not spill over to an offense context that involves a different victim. Similarly, people might not repair their moral self-image after withholding an apology by behaving morally toward a person

other than the victim who was originally harmed by the offense. It is therefore possible that apologizing does not affect subsequent moral behavior toward different targets. Future work could explore the current hypotheses in contexts that involve the same victim and examine potential moderators, such as whether the victim offers forgiveness. If transgressors apologize with the goal of being forgiven, then progressing toward their forgiveness goal might increase the occurrence of licensing (Mullen & Monin, 2016).

Another possible reason for the null effects is that an apology simply balances out the offense, thereby restoring the transgressors' moral self-image but consequently having no impact on their subsequent behavior. The results from Study 2 suggested that apologizing does not necessarily bolster moral self-image above baseline levels, as seemed to be the case in Study 1. Although apologizing has a restorative effect on moral self-image, apologizing might be limited by its inability to then spillover to impact subsequent moral decisions. Indeed, the current studies differ from prior work on moral licensing in that the moral behavior (i.e., the apology) was preceded by an immoral behavior (i.e., the offense), which could be contributing to a weaker effect on subsequent moral behavior. That is, perhaps the offense plus an apology completes a person's moral self-regulation cycle.

If so, future work could investigate various factors that might impact the effect of apologizing and not apologizing on transgressors' moral self-image, such as offense severity, the ambiguity of the subsequent behavior, and apology quality. The results from Study 2 suggest that the perceived severity of the offense might play a role in moral cleansing. However, this interaction pattern was not replicated in Study 3 and therefore remains preliminary. Nevertheless, it seems plausible that the severity of the offense being repaired could impact how transgressors feel and behave after apologizing. Regarding the ambiguity of the subsequent behavior, Mullen and Monin

(2016) posit that moral behavior has greater licensing power when the subsequent behavior is more ambiguous and can be attributed to mitigating external circumstances, which could be relevant in offense contexts that involve the same victim. Regarding apology quality, higher quality apologies might help transgressors feel a stronger boost to their moral self-image. Future work could test this experimentally by assigning participants to a high vs. low quality apology condition to examine whether higher quality apologies have a larger impact on moral self-image and subsequent moral behavior.

The current study also did not find evidence that moral identity or proclivity to apologize moderated the impact of apologizing on moral behavior. Prior work suggests that moral behavior has less licensing power for individuals who identify strongly with the moral issue at hand. For instance, Effron and colleagues (2009) found that participants who established their moral credentials by indicating their preference for Obama (vs. McCain) around the time of the 2008 election were less likely to support allocating government funds to an organization working primarily in a Black community, but not if they scored low on a measure of modern racism. In the current research, perhaps a general moral identity scale was not specific enough to the moral values being restored via an apology to impact licensing effects. Regarding proclivity to apologize, I expected that apologizing could have less licensing power for those who were more dispositionally oriented toward apologizing for their offenses. However, proclivity to apologize was not a significant moderator in Studies 2 and 3. This might suggest that apologies have a similar effect on peoples' moral self-image and tendency to morally license across high and low levels of proclivity to apologize.

Finally, it is worth noting that there have been several failed replication attempts of moral licensing and moral cleansing effects (Blanken et al., 2014; Earp et al., 2014; Rotella & Barclay,

2020). It is difficult to discern whether these failed replications are due to false positives, differences in methodology, or whether there are hidden moderators at play. However, a meta-analysis conducted by Blanken and colleagues (2015) found an overall moral licensing effect of *d* = .30, which lends confidence to the validity of moral licensing and points to the need for more research to determine the conditions and contexts under which it occurs.

Despite the null effects of the apology condition on moral behavior, the results showing the impact that apologizing vs. not apologizing had on moral self-image are encouraging—these changes could potentially spill over and impact subsequent moral behavior. However, it remains unclear whether apologizing might lead to licensing or consistency. To the extent that apologizing leads to licensing, future work investigating this possibility could provide a better understanding of the barriers to moral and prosocial behavior, as well as the processes that could lead to perpetuating continued harm in the same relationship. To the extent that apologizing leads to consistency, this could have positive implications for restorative justice programs. Restorative justice programs seem to reduce recidivism compared to non-restorative or retributive processes in the criminal justice system (Latimer et al., 2005; Rodriguez, 2007). Suzuki and Jenkins (2022) posit that there is a dearth of research on the role of the apology-forgiveness cycle in restorative justice programs, therefore the impact that apologizing has on transgressors' moral self-image in this context is unknown. Although speculative, apologizing could have a restorative effect on transgressors' moral and social identity in a way that might positively impact future behavior.

6.0 Conclusion

Apologizing can repair the adverse effect that an offense has on transgressors' moral self-image, whereas withholding an apology can leave transgressors in a morally depleted state. However, little is known about whether apologizing could morally license transgressors to perpetuate harm or whether apologizing could lead to moral consistency. Future work could explore the conditions under which apologizing motivates licensing versus consistency, such as different vs. same victim contexts, the severity of the offense, the ambiguity of the subsequent behavior, the goal of apologizing and the role of victim forgiveness, and the role of apology quality. Investigating these potential negative or positive downstream consequences of apologies will offer insights into whether there are conditions under which apologies might backfire, freeing transgressors up to withdraw from subsequent opportunities to engage in moral behavior, or whether apologies motivate moral consistency and thus future prosocial behavior.

Appendix A Tables and Figures

Table 1 Means and Standard Deviations for Moral Self-Image Ratings, Study 1

Condition	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 3 *M* (*SD*)
No Apology	4.87_{a} (1.13)	4.21_{b} (1.24)	3.92_{c} (1.25)
Apology	4.79_{a} (1.00)	4.14_{b} (1.18)	5.30_{d} (1.01)

Note. The scale ranged from 1-7.

Table 2 Correlation Matrix of All Dependent Measures, Study 2

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Moral Self-Image	—												
2. Moral Licensing	.05	—											
3. Self-Negativity	-.18**	-.01	—										
4. Other-Negativity	-.26***	0.00	.70***	—									
5. Self-apology deservingness	-.21**	.09	.46***	.74***	—								
6. Other-apology deservingness	-.08	.10	.67***	.55***	.60***	—							
7. Self-Severity	.13	-.05	.51***	.21*	.13	.50***	—						
8. Other-Severity	-.01	-.08	.20*	.57***	.48***	.07	.42***	—					
9. Moral Identity	.23***	.46***	0.00	-.03	.03	.08	-.04	-.09	—				
10. PAM	.26***	.36***	-.08	-.12	0.00	.10	.04	-.17	61***	—			
11. Apologetic	.45***	.17*	-.13	-.09	.14	.17*	.01	-.01	.06	.30***	—		
12. Responsibility	.48***	.21*	-.13	-.07	.14	.17*	-.03	0.00	.09	.25**	.86***	—	
13. Immorality	.04	.42***	.13	.06	.13	.11	.05	-.04	.41***	.20**	.10	.08	—

Note. The correlations for self-severity, other-severity, apologetic, and responsibility are based on data from the apology and no apology conditions because these outcomes were not measured in the control condition. *** $p < .001$. ** $p < .01$. * $p < .05$.

Table 3 Descriptive Statistics and ANOVAs, Study 2

Measure	Condition					
	Control	No Apology	Apology	$F(2, 208)$	p	η^2
	M (SD)	M (SD)	M (SD)			
Moral Self-Image	5.20_a (1.08)	4.02_b (1.26)	4.97_a (1.25)	20.86	< .001	.167
Moral Licensing	5.70_a (1.47)	5.91_a (1.42)	6.11_a (1.12)	1.64	.197	.016
Self-Negativity	1.76_a (1.21)	3.76_b (1.36)	3.40_b (1.50)	43.13	< .001	.293
Other-Negativity	1.73_a (1.25)	4.82_b (1.23)	4.32_b (1.47)	110.04	< .001	.514
Self-apology deservingness	1.70_a (1.27)	4.59_b (1.73)	4.65_b (1.71)	80.27	< .001	.436
Other-apology deservingness	1.93_a (1.51)	4.10_b (1.62)	4.35_b (1.54)	51.89	< .001	.333
Immorality	5.20_a (1.52)	5.38_a (1.72)	5.75_a (1.14)	2.54	.081	.024
				$t(138)$	p	Cohen's d
Self-Severity	—	2.84_a (1.38)	2.89_a (1.24)	.23	.819	.039
Other-Severity	—	4.15_a (1.49)	3.67_a (1.37)	1.99	.049	.34
Apologetic	—	2.93_a (1.76)	5.86_b (1.07)	12.03	< .001	2.03
Accept Responsibility	—	3.49_a (1.94)	6.00_b (1.25)	9.20	< .001	1.56

Note. Each scale ranged from 1-7. Means in the same row with different subscripts are significantly different from each other based on multiple comparisons with Bonferroni correction. Regarding moral licensing, lower means indicate greater moral licensing.

Table 4 Regression Analyses, Study 2

Moderator Predictor	*b*	*SE*	95% CI	*t*	*p*
Moral Identity					
(Constant)	5.91	.08	(5.75, 6.07)	71.64	<.001
Apology vs. no apology & control (D1)	-.19	.18	(-.53, .16)	-1.05	.295
No apology vs. control (D2)	.25	.20	(-.15, .65)	1.23	.219
Moral identity	.72	.09	(.53, .91)	7.57	<.001
D1*Moral identity	.14	.21	(-.27, .53)	.66	.510
D2*Moral identity	-.40	.23	(-.86, .05)	-1.75	.082
PAM					
(Constant)	5.90	.09	(5.73, 6.08)	66.15	<.001
Apology vs. no apology & control (D1)	-.14	.19	(-.52, .23)	-.75	.452
No apology vs. control (D2)	.35	.22	(-.08, .78)	1.61	.110
PAM	.38	.07	(.25, .52)	5.46	<.001
D1*PAM	.03	.15	(-.27, .33)	.19	.851
D2*PAM	-.18	.17	(-.52, .15)	-1.08	.282
Self-Negativity					
(Constant)	5.79	.11	(5.57, 6.01)	51.91	<.001
Apology vs. no apology & control (D1)	-.58	.22	(-1.01, -.14)	-2.60	.010
No apology vs. control (D2)	.40	.29	(-.18, .98)	1.37	.172
Self-negativity	-.08	.07	(-.22, .05)	-1.18	.238
D1*Self-negativity	.08	.14	(-.20, .35)	.56	.574
D2*Self-negativity	.39	.17	(.05, .74)	2.22	.027
Other-Negativity					
(Constant)	5.83	.14	(5.56, 6.10)	42.78	<.001
Apology vs. no apology & control (D1)	-.53	.26	(-1.04, -.03)	-2.07	.039
No apology vs. control (D2)	.60	.37	(-.13, 1.32)	1.62	.107
Other-negativity	-.10	.07	(-.25, .04)	-1.46	.146
D1*Other-negativity	-.01	.14	(-.30, .27)	-.10	.921
D2*Other-negativity	.16	.19	(-.21, .52)	.85	.395
Self-Apology Deservingness					
(Constant)	5.74	.13	(5.48, 5.99)	43.91	<.001
Apology vs. no apology & control (D1)	-.49	.26	(-.99, .01)	-1.92	.056

No apology vs. control (D2)	.50	.35	(-.18, 1.18)	1.45	.149
Self-apology deservingness	-.02	.06	(-.14, .10)	1.83	.776
D1*Self-apology deservingness	-.11	.12	(-.35, .14)	-.86	.394
D2*Self-apology deservingness	.29	.16	(-.02, .60)	1.83	.068
Other-Apology Deservingness					
(Constant)	5.81	.11	(5.59, 6.03)	51.06	<.001
Apology vs. no apology & control (D1)	-.33	.24	(-.80, .13)	-1.42	.157
No apology vs. control (D2)	.29	.29	(-.28, .86)	1.00	.318
Other-apology deservingness	.03	.06	(-.09, .15)	.55	.585
D1*Other-apology deservingness	-.09	.13	(-.34, .17)	-.68	.497
D2*Other-apology deservingness	.20	.15	(-.09, .49)	1.38	.169

Table 5 Correlation Matrix of All Dependent Measures, Study 3

	1	2	3	4	5	6	7	8
1. Moral Self-Image	—							
2. Donations	.11**	—						
3. Negativity	-.17***	-.05	—					
4. Closeness	.09*	.07	.04	—				
5. Moral Identity	.30***	.15***	-.07	.07	—			
6. PAM	.31***	.15***	-.10*	-.01	.63***	—		
7. Comprehensiveness	.06	.12	.03	.14	.07	0.00	—	
8. Defensiveness	0.00	-.09	-.05	-.05	-.14	-.03	-.25*	—

Note. The correlations for comprehensiveness and defensiveness are based on data from the apology condition. *** $p < .001$. ** $p < .01$. * $p < .05$.

Table 6 Descriptive Statistics and ANOVAs, Study 3

Measure	Control *M* (*SD*)	No Apology *M* (*SD*)	Apology *M* (*SD*)	*F*(2, 590)	*p*	η^2
Moral Self-Image	5.21_a (.99)	4.72_b (1.26)	4.84_b (1.12)	11.30	< .001	.037
Donations	$.31_a$ (.38)	$.26_a$ (.37)	$.30_a$ (.38)	.93	.395	.003
Negativity	1.64_a (1.18)	4.97_b (1.31)	4.88_b (1.35)	457.27	< .001	.608
Closeness	5.03_a (1.77)	5.04_a (1.78)	5.20_a (1.75)	.53	.588	.002

Note. Each scale ranged from 1-7. Donations ranged from 0 cents to $1.00, with lower mean donations indicating greater moral licensing. Means with different superscripts in the same row are significantly different from each other based on multiple comparisons with Bonferroni correction.

Table 7 Regression Analyses, Study 3

Moderator Predictor	*b*	*SE*	95% CI	*t*	*p*
Moral Identity					
(Constant)	.30	.02	(.27, .34)	15.54	< .001
Apology vs. no apology & control (D1)	-.01	.03	(-.08, .06)	-.21	.833
No apology vs. control (D2)	-.04	.04	(-.11, .03)	-1.01	.315
Moral identity	.06	.02	(.02, .10)	2.74	.006
D1*Moral identity	-.03	.04	(-.12, .05)	-.79	.429
D2*Moral identity	-.02	.02	(-.05, .01)	-1.18	.237
PAM					
(Constant)	.29	.02	(.26, .32)	18.83	<.001
Apology vs. no apology & control (D1)	-.01	.04	(-.08, .06)	-.37	.711
No apology vs. control (D2)	-.03	.04	(-.10, .04)	-.92	.356
PAM	.04	.02	(-.01, .08)	1.53	.126
D1*PAM	-.01	.03	(-.07, .05)	-.43	.666
D2*PAM	.01	.03	(-.05, .07)	.23	.820
Closeness					
(Constant)	.29	.02	(.26, .32)	18.77	<.001
Apology vs. no apology & control (D1)	-.01	.04	(-.08, .06)	-.27	.791
No apology vs. control (D2)	-.05	.04	(-.12, .02)	-1.32	.186
Closeness	.01	.01	(-.01, .04)	.97	.334
D1* Closeness	0.00	.02	(-.04, .04)	-.16	.875
D2* Closeness	0.00	.02	(-.04, .04)	-.06	.952
Negativity					
(Constant)	.28	.02	(.24, .33)	11.63	<.001
Apology vs. no apology & control (D1)	.01	.05	(-.09, .10)	.10	.919
No apology vs. control (D2)	.01	.06	(-.12, .13)	.11	.914
Negativity	-.01	.02	(-.06, .03)	-.61	.541
D1* Negativity	-.04	.03	(-.09, .02)	-1.34	.182
D2* Negativity	-.01	.03	(-.07, .05)	-.31	.754

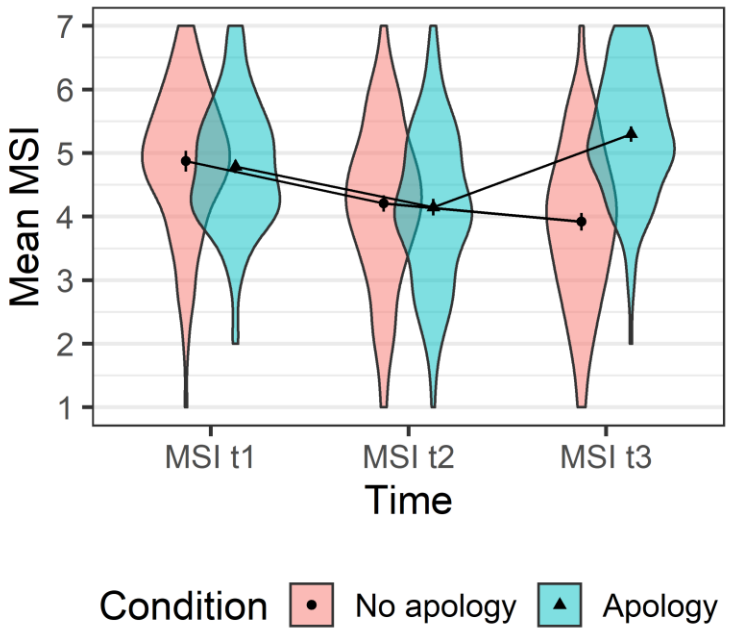

Figure 1 Changes in moral self-image post-offense and post-apology by condition, Study 1

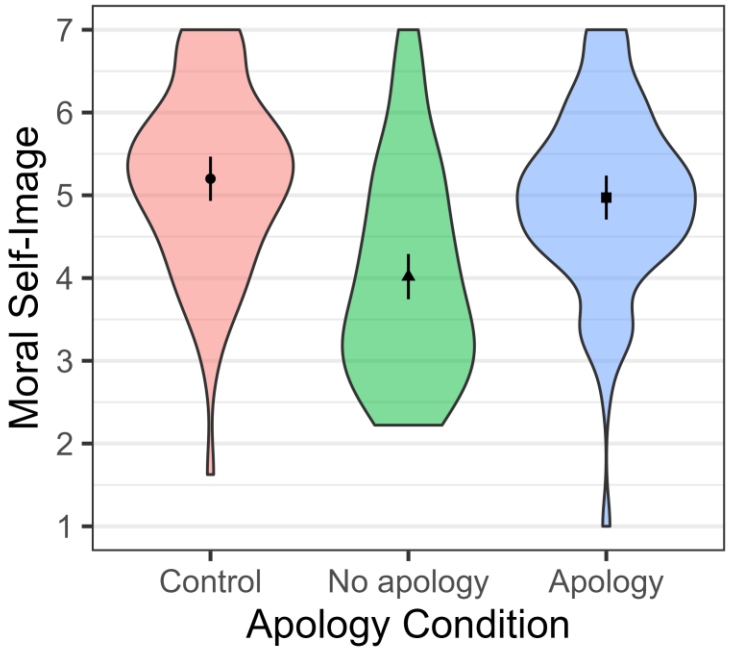

Figure 2 The effect of apology condition on moral self-image, Study 2

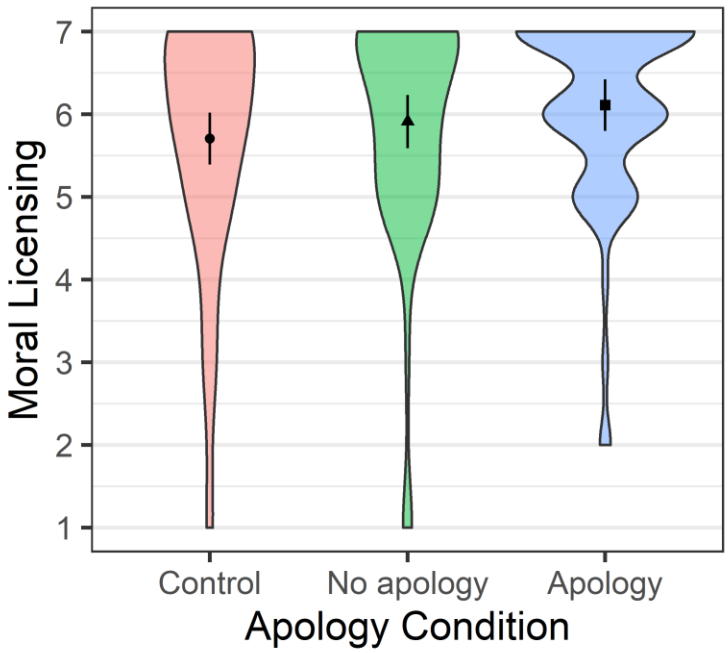

Figure 3 The effect of apology condition on moral behavior, Study 2

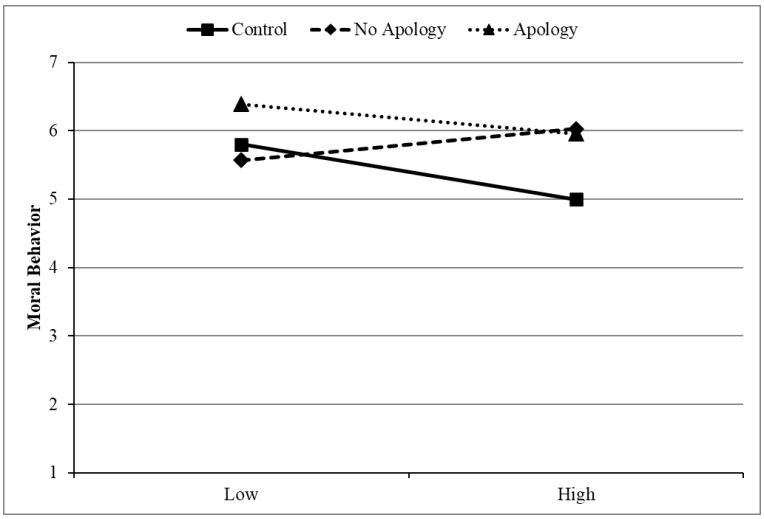

Figure 4 Interaction between self-negativity and apology condition on moral behavior, Study 2[7]

[7] *Note*: Lower moral behavior indicates greater moral licensing.

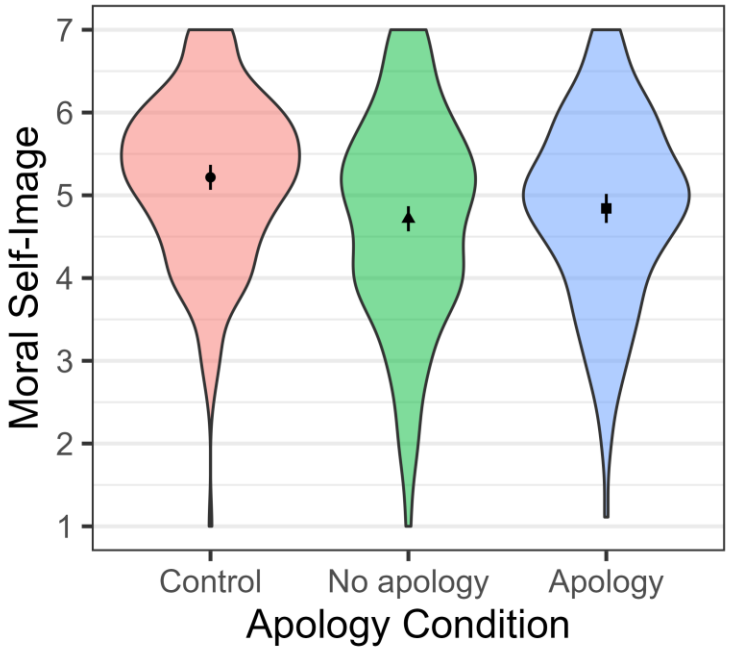

Figure 5 The effect of apology condition on moral self-image, Study 3

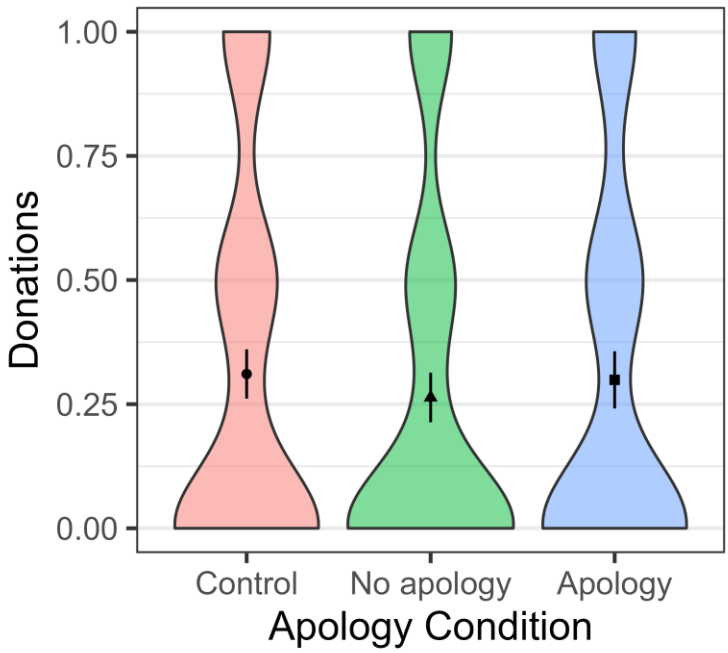

Figure 6 The effect of apology condition on donations, Study 3

Appendix B Study 1 Hypothetical Offense

Participants in both conditions (apology and no apology) will read the following scenario:

"*It's Sunday morning and you're having breakfast with your roommate, Taylor. During your conversation, Taylor comments on the messiness of the apartment—there are dirty dishes in the sink, the floors need to be cleaned in the kitchen and living room, and the shared bathroom needs to be cleaned. To address this issue, Taylor proposes that both of you stick to a regular cleaning schedule. Taylor suggests that they clean the living room and bathroom and that you clean the kitchen. You agree to this division of work and promise to do a deep clean of the kitchen after work on Monday because you already have weekend plans and prior commitments.*

That Monday, you get home from work and see that Taylor unexpectedly has several friends over. You notice that Taylor cleaned up the living room and bathroom, but it seems like a bad time to clean the kitchen because it's late and you're tired and Taylor has friends over. Instead of cleaning the kitchen as you had promised, you decide to spend the rest of the evening relaxing in your room while Taylor hangs out with friends. Taylor's hangout session turns out to be more of a small party, which unfortunately lasts until 1:00am in the morning. You wonder why Taylor didn't let you know about having people over, which is something the both of you had previously agreed to try to do for each other. Everyone was very loud, keeping you up and making it difficult to get any sleep. Taylor knows that you have work early the next morning but did not seem to consider how this might affect you.

You wake up for work the next day and make a cup of coffee. However, you end up going into work feeling tired and sluggish after getting so little sleep that night."

Appendix C Study 1 Experimental Conditions

Apology condition:
Participants in this condition will continue to read the scenario and imagine themselves texting their roommate an apology:

Now, continuing the story from before…

…During your break at work, you spend some time mulling over the events that transpired yesterday. There's clearly tension between you and Taylor that's making things uncomfortable. You realize that you both did things to upset each other—you didn't clean the kitchen and Taylor had people over late without letting you know. You decide to offer Taylor a heartfelt apology for your part in the situation. You won't see Taylor until later tonight because they always work late on Tuesdays, so you decide to send an apology by text, saying:

'Hey, I'm sorry that I didn't clean the kitchen yesterday even though I promised I would. I was really tired from work, but I should have at least texted you to explain. I want you to know that from now on I'll try my best to stick to my cleaning schedule because I know it's really important to you."

No apology condition:
Participants in this condition will continue to read the scenario and decide to not offer their roommate an apology:

Now, continuing the story from before…

…During your break at work, you spend some time mulling over the events that transpired yesterday. There's clearly tension between you and Taylor that's making things uncomfortable. You realize that you both did things to upset each other—you didn't clean the kitchen and Taylor had people over late without letting you know. You consider sending a text to Taylor to apologize for breaking your promise to clean the kitchen, but in the end, you decide not to do so.

Appendix D Moral Self-Image Scale

Instructions:

Time point	Preamble
t1	Instructions: Please respond to the following statements as they apply to you.
t2	Instructions: We would now like you to complete the same questions again, based on the scenario that you just imagined yourself in. Please respond to the following statements as they apply to you.
t3	Instructions: We would now like you to complete the same questions one last time, based on the second part of the scenario you imagined yourself in. Please respond to the following statements as they apply to you.

Moral Self-Image Scale:

1. Compared to the caring person I want to be, right now I feel I am:

Not at all caring						*Extremely caring*
1	2	3	4	5	6	7

2. Compared to the compassionate person I want to be, right now I feel I am:

Not at all compassionate						*Extremely compassionate*
1	2	3	4	5	6	7

3. Compared to the fair person I want to be, right now I feel I am:

Not at all fair						*Extremely fair*
1	2	3	4	5	6	7

4. Compared to the friendly person I want to be, right now I feel I am:

Not at all friendly						*Extremely friendly*
1	2	3	4	5	6	7

5. Compared to the generous person I want to be, right now I feel I am:

Not at all generous						*Extremely generous*
1	2	3	4	5	6	7

6. Compared to the moral person I want to be, right now I feel I am:

Not at all moral						*Extremely moral*
1	2	3	4	5	6	7

7. Compared to the helpful person I want to be, right now I feel I am:

Not at all helpful						*Extremely helpful*
1	2	3	4	5	6	7

8. Compared to the honest person I want to be, right now I feel I am:

Not at all honest						*Extremely honest*
1	2	3	4	5	6	7

9. Compared to the kind person I want to be, right now I feel I am:

Not at all kind						*Extremely kind*
1	2	3	4	5	6	7

Appendix E Study 1 Measures

Based on the scenario with Taylor, to what extent…(1 = *Not at all,* 7 = *Very*)
…did you offend Taylor?
…did Taylor offend you?
...how severe is what you did to Taylor?
...how severe is what Taylor did to you?
...how responsible are you for any bad feelings between you and Taylor?...how responsible is Taylor for any bad feelings between you and Taylor?...how justifiable were your actions toward Taylor?
...how justifiable were Taylor's actions toward you?
...to what extent did Taylor deserve an apology from you?...to what extent did you deserve an apology from Taylor?
How likely would you be to apologize in this situation? (*1 = Not at all likely, 7 = Extremely likely*)

How moral would it be to apologize in this situation? (*1 = Not at all moral, 7 = Extremely moral*)

Appendix F Study 2 Experimental Conditions

Participants in both conditions (apology and no apology) will read the following scenario:

"*It's Sunday morning and you're having breakfast with your roommate, Taylor. During your conversation, Taylor comments on the messiness of the apartment—there are dirty dishes in the sink, the floors need to be cleaned in the kitchen and living room, and the shared bathroom needs to be cleaned. To address this issue, Taylor proposes that both of you stick to a regular cleaning schedule. Taylor suggests that they clean the living room and bathroom and that you clean the kitchen. You agree to this division of work and promise to do a deep clean of the kitchen after work on Monday because you already have weekend plans and prior commitments.*

That Monday, you get home from work and see that Taylor unexpectedly has several friends over. You notice that Taylor cleaned up the living room and bathroom, but it seems like a bad time to clean the kitchen because it's late and you're tired and Taylor has friends over. Instead of cleaning the kitchen as you had promised, you decide to spend the rest of the evening relaxing in your room while Taylor hangs out with friends. Taylor's hangout session turns out to be more of a small party, which unfortunately lasts until 1:00am in the morning. You wonder why Taylor didn't let you know about having people over, which is something the both of you had previously agreed to try to do for each other. Everyone was very loud, keeping you up and making it difficult to get any sleep. Taylor knows that you have work early the next morning but did not seem to consider how this might affect you.

You wake up for work the next day and make a cup of coffee. However, you end up going into work feeling tired and sluggish after getting so little sleep that night."

Apology condition:
During your break at work, you spend some time mulling over the events that transpired yesterday. There's clearly tension between you and Taylor that's making things uncomfortable. You realize that you both did things to upset each other—you didn't clean the kitchen and Taylor had people over late without letting you know. You decide to offer Taylor a heartfelt apology for your part in the situation. You won't see Taylor until later tonight because they always work late on Tuesdays, so you decide to send an apology by text, saying:

'Hey, I'm sorry that I didn't clean the kitchen yesterday even though I promised I would. I was really tired from work, but I should have at least texted you to explain. I want you to know that from now on I'll try my best to stick to my cleaning schedule because I know it's really important to you."

No apology condition:
During your break at work, you spend some time mulling over the events that transpired yesterday. There's clearly tension between you and Taylor that's making things uncomfortable. You realize that you both did things to upset each other—you didn't clean the kitchen and Taylor had people

over late without letting you know. You consider sending a text to Taylor to apologize for breaking your promise to clean the kitchen, but in the end, you decide not to do so.

No offense control condition:
It's Sunday morning and you're having breakfast with your roommate, Taylor. During your conversation, Taylor comments on the messiness of the apartment—there are dirty dishes in the sink, the floors need to be cleaned in the kitchen and living room, and the shared bathroom needs to be cleaned. To address this issue, Taylor proposes that both of you stick to a regular cleaning schedule. Taylor suggests that they clean the living room and bathroom and that you clean the kitchen. You agree to this division of work and promise to do a deep clean of the kitchen after work on Monday because you already have weekend plans and prior commitments.

That Monday, you get home late and see that Taylor has several friends over. You notice that Taylor cleaned up the living room and bathroom, but it seems like a bad time to clean the kitchen because it's late and you're tired and Taylor has friends over. However, you decide clean the kitchen anyways to fulfill your end of the agreement.
Afterwards, you spend the rest of the evening relaxing in your room. Taylor's friends leave around 10:00pm, and you get a good night's sleep.

You wake up for work the next day and make a cup of coffee. You go to work feeling prepared for your meetings and the day ahead.
During your break at work, you spend some time mulling over the cleaning schedule that Taylor proposed. It seems like it's working great because your apartment looks nice, and you both can contribute equally to the cleaning.

Appendix G Study 2 Workplace Scenario and Main Dependent Variable

A little later that day, you're at work and busy finishing an important task. Your supervisor recently assigned you to lead a project and assigned your coworker, named Jamie, to assist on the project. You have been working hard on this project all week and have an upcoming deadline to submit a report to your supervisor.

However, you're at a crossroad with the project and are stuck trying to solve a particular problem. You decide to call Jamie to ask for advice because they have experience working on similar tasks. You and Jamie begin brainstorming ideas, and Jamie comes up with an innovative solution. After the meeting, you integrate Jamie's solution by further developing it and including it as a part of the report.

Later, you show your supervisor the finished report and hope that they are impressed with the work. Your yearly review is coming up and you might be able to get a promotion or pay-raise, because of the importance of this particular task.

Before the end of the day, your supervisor stops by your office. Your supervisor tells you that the report looks good overall. Your supervisor then specifically mentions the problem that you had been stuck on, and says that they found your solutions to that problem really impressive and creative. Because you were the lead on the project, your supervisor is unaware that some of the ideas in the report were actually based on Jamie's ideas even though they were assigned to assist on the project. Before your supervisor walks away, you have the opportunity to clarify the role that Jamie played in coming up with the solutions that your supervisor specifically praised. You know this conversation will stay between you and your supervisor.

Main Dependent Variable (Index of moral licensing and moral cleansing):
How likely would you be to clarify Jamie's contributions to the task? (*1 = Not at all likely, 7 = Extremely likely*)

Other measures:
How immoral would it be to take credit for Jamie's contributions to the task? (*1 = Not at all immoral, 7 = Extremely immoral*)[8]

[8] The theoretical framework is that transgressors can earn moral credits through offering high-quality apologies. An alternative mechanism involves moral credentials, where apologies allow transgressors to reconstrue their subsequent behavior as more moral. The prediction is that if moral licensing results from moral credits, then moral licensing should occur while controlling for the perceived immorality of the subsequent behavior. If, however, moral licensing occurs because of moral credentials, then moral licensing should occur through lower perceptions of immorality.

Understanding whether the mechanism driving moral licensing effects is not a primary goal of the proposed work, but I am including this item to explore these two possible mechanisms.

Appendix H Study 2 Measures

Severity:
To what extent did you behave negatively toward Taylor in this situation? (*1 = Not at all negative, 7 = Extremely negative*)

To what extent did Taylor behave negatively toward you in this situation? (*1 = Not at all negative, 7 = Very negatively*)

Apology and No Apology condition only:
How severe is what Taylor did to you? (*1 = Not at all severe, 7 = Very severe*)
How severe is what you did to Taylor? (*1 = Not at all severe, 7 = Very severe*)

Apology deservingness:
To what extent…
…did Taylor deserve an apology? (*1 = To a very small extent, 7 = To a very large extent*)
…did you deserve an apology from Taylor? (*1 = To a very small extent, 7 = To a very large extent*)

Manipulation check questions for the offense conditions:
Based on how you responded to Taylor in the story, to what extent…
…were you apologetic? (*1 = Not at all apologetic, 7 = Extremely apologetic*)
…did you accept responsibility for your actions? (*1 = To a very small extent, 7 = To a very large extent)*

Appendix I Study 2 Exploratory Moderators

Proclivity to Apologize Measure: (*1 = Strongly Disagree, 7 = Strongly Agree*)
1. I tend to downplay my wrongdoings to the other person, rather than apologize
2. I tend not to apologize because I could get into trouble for confessing
3. If I think no one will know what I have done, I am likely not to apologize
4. By not apologizing, it allows me to continue to behave as I want to behave
5. To avoid feeling incompetent, I tend not to apologize
6. I don't apologize very often because I don't like to admit that I'm wrong
7. I don't like to apologize because it lets the other person feel superior to me
8. m

Moral Identity Questionnaire: (*1 = Strongly Disagree, 7 = Strongly Agree*)
1. I try hard to act honestly in most things I do.
2. Not hurting other people is one of the rules I live by.
3. It is important for me to treat other people fairly.
4. I want other people to know they can rely on me.
5. I always act in ways that do the most good and least harm to other people.
6. If doing something will hurt another person, I try to avoid it even if no one would know.
7. One of the most important things in life is to do what you know is right.
8. Once I′ve made up my mind about what is the right thing to do, I make sure I do it.
9. As long as I make a decision to do something that helps me, it does not matter much if other people are harmed.
10. It is ok to do something you know is wrong if the rewards for doing it are great.
11. If no one is watching or will know it does not matter if I do the right thing.
12. It is more important that people think you are honest than being honest.
13. If no one could find out, it is okay to steal a small amount of money or other things that no one will miss.
14. There is no point in going out of my way to do something good if no one is around to appreciate it.
15. If a cashier accidentally gives me $10 extra change, I usually act as if I did not notice it.
16. Lying and cheating are just things you have to do in this world.
17. Doing things that some people might view as not honest does not bother me.
18. If people treat me badly, I will treat them in the same manner.
19. I will go along with a group decision, even if I know it is morally wrong.
20. Having moral values is worthless in today's society.

Appendix J Study 3 Conditions and Measures

Recall Prompt:
We would like you to think about something that you have done to hurt, upset, or offend someone. Please think of an offense that is **currently unresolved – something that you did that is unsettled or an ongoing issue**. Sometimes it takes a few minutes to think of something that you did to offend someone. If an offense doesn't come to mind right away, please take some time to think about it. Once you have an offense in mind, please go to the next page of the survey.

Please provide a short description of the offense:_____

Apology Condition:
We would now like you to write a letter to this person. There are many things you could write in this letter. However, we would like to make a special request that you use this letter as an opportunity to apologize. If you choose to follow this request, it is completely up to you what you say to in your apology.

No Apology Condition:
We would now like you to think about the last interaction you had with someone—someone other than the person you just thought about regarding an unresolved conflict. This could be any kind of interaction, for example, the last conversation you had with someone or the last activity you did with someone. Take a few minutes to think about the interaction once you have it in mind.

Control Condition:
We would like you to think about the last interaction you had with someone. Please think about the last time you interacted with someone—this could be any kind of interaction, for example, the last conversation you had with someone or the last activity you did with someone. Once you have an interaction in mind, please go to the next page of the survey.

Measures:
Closeness:
How close were you to this person at the time of the interaction? (*1 = Not at all close, 7 = Extremely close*)

Negativity:
How negative was the interaction? (*1 = Not at all negative, 7 = Extremely negative*)
How upsetting was the interaction? (*1 = Not at all upsetting, 7 = Extremely upsetting*)
How offensive was the interaction? (1= Not at all offensive, 7 = Extremely offensive) (severity item)
How friendly was the interaction? (1 =*Not at all friendly*, 7=*Extremely friendly*) (filler item)
How supportive was the interaction? (1=*Not at all supportive*, 7=*Extremely supportive*) (filler item)
How joyful was the interaction? (1=Not at all joyful, 7=Extremely joyful) (filler item)

Apology and No Apology Conditions only:
How severe was the offense? (*1 = Not at all severe, 7 = Very severe*)

Appendix K Self-Rated Apology Quality

Instructions: Please complete the following questions based on what you wrote in your apology letter.

Please respond to the following statements using the scale provided, from 1 (Strongly disagree) to 7 (Strongly agree).

1. I apologized or expressed remorse about what happened.
2. I accepted responsibility for what happened.
3. I offered to make up for the damage that was caused.
4. I explained my behavior without pushing the blame onto external factors.
5. I promised to behave better from now on.
6. I acknowledged that they suffered or they have been inconvenienced by what happened.
7. I admitted that my actions were wrong or unfair.
8. I asked them for forgiveness.
9. I justified my behavior.
10. I put some of the blame for what happened on them.
11. I gave excuses for what happened.
12. I minimized the severity of what happened.

Appendix L Study 3 Donation Scenario, Main Dependent Variable, and Suspicion Check

Donation:
Thank you for your participation on this study. As described at the beginning of the study, you will be paid $2.50 for your participation. In addition, due to additional funds that we received for this research, we will be giving you a bonus of $1.00. You may keep this money for yourself, or, if you like, you can donate some or all of the bonus money to GiveWell's Maximum Impact Fund, which is a non-profit organization that donates to charities that save or improve lives the most per dollar. This decision is completely up to you: You can donate none, all, or some portion of this bonus.

How much would you like to donate? (*0 cents – $1.00*)

Suspicion check:
What do you believe the purpose of the study is? (Please describe):

Bibliography

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, 6, 1–62. Academic Press.

Black, J. E., & Reynolds, W. M. (2016). Development, reliability, and validity of the Moral Identity Questionnaire. *Personality and Individual Differences*, *97*, 120–129. https://doi.org/10.1016/j.paid.2016.03.041

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality & social psychology bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

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

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

Darby, B. W., & Schlenker, B. R. (1982). Children's reactions to apologies. *Journal of Personality and Social Psychology*, *43*(4), 742–753. https://doi.org/10.1037/0022-3514.43.4.742

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

Earp, B. D., Everett, J. A., Madva, E. N., & Hamlin, J. K. (2014). Out, damned spot: Can the "Macbeth Effect" be replicated? *Basic and Applied Social Psychology*, *36*(1), 91-98. https://doi.org/10.1080/01973533.2013.856792

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

Exline, J. J., Deshea, L., & Holeman, V. T. (2007). Is apology worth the risk? Predictors, outcomes, and ways to avoid regret. *Journal of Social and Clinical Psychology*, *26*(4), 479–504. https://doi.org/10.1521/jscp.2007.26.4.479

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*(5), 894–914. https://doi.org/10.1037/a0019993

Fisher, M. L., & Exline, J. J. (2010). Moving toward self-forgiveness: Removing barriers related to shame, guilt, and regret. *Social and Personality Psychology Compass*, *4*(8), 548–558. https://doi.org/10.1111/j.1751-9004.2010.00276.x

Forster, D. E., Billingsley, J., Burnette, J. L., Lieberman, D., Ohtsubo, Y., & McCullough, M. E. (2021). Experimental evidence that apologies promote forgiveness by communicating relationship value. *Scientific Reports*, *11*(1), 13107. https://doi.org/10.1038/s41598-021-92373-y

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. http://www.jstor.org/stable/41406379

Hall, J. H., & Fincham, F. D. (2005). Self–Forgiveness: The stepchild of forgiveness research. *Journal of Social and Clinical Psychology*, *24*(5), 621–637. https://doi.org/10.1521/jscp.2005.24.5.621

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Hechler, S., Wenzel, M., Woodyatt, L., & de Vel-Palumbo, M. (2022). What does being hard on yourself communicate to others? The role of symbolic implications of self-punishment in attributions of remorse. *Journal of Experimental Social Psychology*, *101*, 104305. https://doi.org/10.1016/j.jesp.2022.104305

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

Howell, A. J., Dopko, R. L., Turowski, J. B., & Buro, K. (2011). The disposition to apologize. *Personality and Individual Differences*, *51*(4), 509–514. https://doi.org/10.1016/j.paid.2011.05.009

Howell, A. J., Turowski, J. B., & Buro, K. (2012). Guilt, empathy, and apology. *Personality and Individual Differences*, *53*(7), 917–922. https://doi.org/10.1016/j.paid.2012.06.021

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*. https://doi.org/10.3389/fpsyg.2015.01878

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

Kiecolt-Glaser, J. K., & Newton, T. L. (2001). Marriage and health: His and hers. *Psychological Bulletin*, *127*(4), 472–503. https://doi.org/10.1037/0033-2909.127.4.472

Kirchhoff, J., Wagner, U., & Strack, M. (2012). Apologies: Words of magic? The role of verbal components, anger reduction, and offence severity. *Peace and Conflict: Journal of Peace Psychology*, *18*(2), 109–130. https://doi.org/10.1037/a0028092

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592095150. https://doi.org/10.1177/2515245920951503

Latimer, J., Dowden, C., & Muise, D. (2005). The Effectiveness of Restorative Justice Practices: A Meta-Analysis. *The Prison Journal*, *85*(2), 127–144. https://doi.org/10.1177/0032885505276969

Leunissen, J. M., De Cremer, D., van Dijke, M., & Reinders Folmer, C. P. (2014). Forecasting errors in the averseness of apologizing. *Social Justice Research*, *27*(3), 322–339. https://doi.org/10.1007/s11211-014-0216-4

Leunissen, J. M., Sedikides, C., & Wildschut, T. (2017). Why narcissists are unwilling to apologize: The role of empathy and guilt. *European Journal of Personality*, *31*(4), 385–403. https://doi.org/10.1002/per.2110

McCullough, M. E., Pedersen, E. J., Tabak, B. A., & Carter, E. C. (2014). Conciliatory gestures promote forgiveness and reduce anger in humans. *Proceedings of the National Academy of Sciences*, *111*(30), 11211–11216. https://doi.org/10.1073/pnas.1405072111

McCullough, M. E., Worthington, E. L., Jr, & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of personality and social psychology*, *73*(2), 321–336. https://doi.org/10.1037//0022-3514.73.2.321

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

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology*, *67*(1), 363–385. https://doi.org/10.1146/annurev-psych-010213-115120

Okimoto, T. G., Wenzel, M., & Hedrick, K. (2013). Refusing to apologize can have psychological benefits (and we issue no mea culpa for this research finding): Refusing to apologize. *European Journal of Social Psychology*, *43*(1), 22–31. https://doi.org/10.1002/ejsp.1901

Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. *Crime & Delinquency*, *53*(3), 355–379. https://doi.org/10.1177/0011128705285983

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

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

Schumann, K. (2018). The psychology of offering an apology: Understanding the barriers to apologizing and how to overcome them. *Current Directions in Psychological Science, 27*(2), 74–78. https://doi.org/10.1177/0963721417741709

Schumann, K. (2014). An affirmed self and a better apology: The effect of self-affirmation on transgressors' responses to victims. *Journal of Experimental Social Psychology, 54*, 89–96. https://doi.org/10.1016/j.jesp.2014.04.013

Schumann, K., & Dragotta, A. (2021). Empathy as a predictor of high-quality interpersonal apologies. *European Journal of Social Psychology*, ejsp.2786. https://doi.org/10.1002/ejsp.2786

Schumann, K., & Walton, G. M. (2022). Rehumanizing the self after victimization: The roles of forgiveness versus revenge. *Journal of Personality and Social Psychology, 122*(3), 469–492. https://doi.org/10.1037/pspi0000367

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology*, *94*(1), 116–132. https://doi.org/10.1037/0022-3514.94.1.116

Stearns, D. C., & Parrott, W. G. (2012). When feeling bad makes you look good: Guilt, shame, and person perception. *Cognition and Emotion*, *26*(3), 407–430. https://doi.org/10.1080/02699931.2012.675879

Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, *64*(6), 885–896. https://doi.org/10.1037/0022-3514.64.6.885

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Suzuki, M., & Jenkins, T. (2022). The role of (self-)forgiveness in restorative justice: Linking restorative justice to desistance. *European Journal of Criminology*, *19*(2), 202–219. https://doi.org/10.1177/1477370819895959

Thai, M., Wenzel, M., & Okimoto, T. G. (2021). Turning tables: Offenders feel like "victims" when victims withhold forgiveness. *Personality and Social Psychology Bulletin*,

014616722110624. https://doi.org/10.1177/01461672211062401

Witvliet, C. V. O., Wade, N. G., Worthington, E. L., Root Luna, L., Van Tongeren, D. R., Berry, J. W., & Tsang, J.-A. (2020). Apology and restitution: Offender accountability responses influence victim empathy and forgiveness. *Journal of Psychology and Theology*, *48*(2), 88–104. https://doi.org/10.1177/0091647120915181

Witvliet, C. V. O., Ludwig, T. E., & Laan, K. L. V. (2001). Granting forgiveness or harboring grudges: Implications for emotion, physiology, and health. *Psychological Science*, *12*(2), 117–123. https://doi.org/10.1111/1467-9280.00320

Woodyatt, L., & Wenzel, M. (2013). Self-Forgiveness and restoration of an offender following an interpersonal transgression. *Journal of Social and Clinical Psychology*, *32*(2), 225–259. https://doi.org/10.1521/jscp.2013.32.2.225

Woodyatt, L., & Wenzel, M. (2014). A needs-based perspective on self-forgiveness: Addressing threat to moral identity as a means of encouraging interpersonal and intrapersonal restoration. *Journal of Experimental Social Psychology*, *50*, 125–135. https://doi.org/10.1016/j.jesp.2013.09.012

Woodyatt, L., Wenzel, M., Okimoto, T. G., & Thai, M. (2022). Interpersonal transgressions and psychological loss: Understanding moral repair as dyadic, reciprocal, and interactionist. *Current Opinion in Psychology*, *44*, 7–11. https://doi.org/10.1016/j.copsyc.2021.08.018

Zechmeister, J. S., Garcia, S., Romero, C., & Vas, S. N. (2004). Don't apologize unless you mean it: A laboratory investigation of forgiveness and retaliation. *Journal of Social and Clinical Psychology*, *23*(4), 532–564. https://doi.org/10.1521/jscp.23.4.532.40309

Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*(5792), 1451–1452. https://doi.org/10.1126/science.1130726