Journal of Experimental Social Psychology 85 (2019) 103874

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Tribalism can corrupt: Why people denounce or protect immoral group members[☆]

Ashwini Ashokkumar[a,*], Meredith Galaif[b], William B. Swann Jr[a]

[a] Department of Psychology, University of Texas Austin, 1 University Station, 108 E. Dean Keeton, Austin, TX 78712-0187, United States of America
[b] EverlyWell, Inc., 800 W Cesar Chavez St, Austin, TX 78701, United States of America

ARTICLE INFO

Keywords:
Moral violations
Ingroup moral transgressions
Group reputation
Moral reputation
Group responsibility
Identity fusion

ABSTRACT

When ingroup members behave immorally, what determines whether other group members denounce vs. protect them? We asked if concerns for group reputation might determine how people respond to the moral indiscretions of group members. In four studies, participants read about an immoral act committed by a member of their political party. The act was either publicly known to people outside of the participant's political party (i.e., "public transgression") or hidden from public view (i.e., "private transgression"). In the public transgression condition, participants endorsed having their political party openly denounce the transgressor in an apparent effort to prevent the party's reputation from being tarnished by association. In the private transgression condition, however, they were reluctant to publicly report the transgressive party member, presumably to prevent reputational loss. Feelings of responsibility for the group arising from either identity fusion with the party or being experimentally assigned to occupy a position of responsibility for the party amplified these effects. Strongly fused participants were even willing to contemplate extreme, unethical actions aimed at protecting party reputation (e.g. tampering with incriminating evidence), regardless of the publicness of the transgression. We conclude that feelings of responsibility for the group and reputational considerations determine whether people denounce or protect ingroup transgressors.

As the #MeToo movement spread, individuals within the entertainment and sports industries denounced transgressive members soon after their appalling moral violations became public (Bleznak, 2018; Maeson & Hobson, 2017). Such denunciations appeared less honorable when it was later revealed that some of these individuals had known about their members' moral violations well before they became public (Maeson & Hobson, 2017). Some of them even went so far as tampering with evidence (Chavez & Sutton, 2018) or coercing victims to sign non-disclosure agreements (Buschmann, Henrichs, Pfeil, Windmann, & Wulzinger, 2017; Watt, 2018). These incidents suggest that group members sometimes go to great lengths to protect the reputation of the group or tribe. When a moral violation within the group is publicly known to outsiders, reputational concerns should compel group members to vociferously denounce the immoral group members in an effort to demonstrate the group's morality. In contrast, when the violation is largely unknown to outsiders, such tribal instincts should compel group members to keep that violation hidden. Group members who feel responsible for protecting the group's reputation should be especially likely to display these effects. Previous research on moral transgressions provides some of the evidentiary basis for these predictions.

1. Moral transgressions and group reputation: the role of the transgression's publicness

Group members want to perceive their group as moral (Ellemers & van den Bos, 2012) and also have others recognize it as such (Ellemers, Pagliaro, Barreto, & Leach, 2008). They accordingly strive to adhere to certain standards to maintain their group's moral self-regard (Pagliaro, Ellemers, & Barreto, 2011) and its moral image (Marques, Paez, & Abrams, 1998; Pagliaro, Ellemers, Barreto, & Di Cesare, 2016). Evidence that their fellow group members have engaged in immoral behavior can undermine these goals by threatening their group's moral self-concept (Van der Toorn, Ellemers, & Doosje, 2015) and moral reputation (Van der Toorn et al., 2015). Individuals who are fused (Buhrmester, 2013) or identified (Biernat, Vescio, & Billings, 1999)

[☆] This paper has been recommended for acceptance by Ernestine Gordijn.

[*] Corresponding author at: Department of Psychology, University of Texas at Austin, 1 University Station, 108 E. Dean Keeton, Austin, TX 78712-0187, United States.

E-mail address: ashwinia@utexas.edu (A. Ashokkumar).

https://doi.org/10.1016/j.jesp.2019.103874
Received 2 April 2019; Received in revised form 27 July 2019; Accepted 30 July 2019

with their group will be especially disturbed by the moral violations of fellow group members. Subsequently, they might selectively disengage from moral cognitions while evaluating the moral transgressions of fellow group members ("Moral Disengagement Theory"; Aquino, Reed II, Thau, & Freeman, 2007; Bandura, 2002), leading them to deny or condone these transgressions (Hofmann, Brandt, Wisneski, Rockenbach, & Skitka, 2018; Van der Toorn et al., 2015), especially when responding to ingroup transgressions that benefit the group (Aguiar, Campos, Pinto, & Marques, 2017). Findings from other research, however, have shown that people denounce immoral group members (Buhrmester, 2013), derogating them even more enthusiastically than immoral outgroup members. This "Black Sheep Effect" (Marques & Yzerbyt, 1988) is understood to reflect an effort to affirm the group's values (Marques, Abrams, Paez, & Martinez-Taboada, 1998).

Simply put, the most prominent theories in the literature make unique predictions regarding whether people denounce or protect transgressive ingroup members, with only few studies providing insight into the mechanisms determining which response people opt for (Aguiar et al., 2017). The current research adds to this body of work by highlighting the role of the publicness of transgressions and reputational concerns in determining when people denounce versus protect ingroup transgressors.

We reason that when people encounter an ingroup moral transgression that is publicly known to people outside the group, they may be motivated to publicly signal their group's morality. Toward this end, they may want their group to publicly denounce the transgressive group member or publicly report them to a relevant external authority (e.g., federal agents). Doing so would help to symbolically distance the group from the transgression, thereby deflecting blame away from the group (Van der Toorn et al., 2015) and publicly affirming the group's moral reputation. On the other hand, when they encounter ingroup moral transgressions that are unknown to people outside their group, the violation will not directly challenge their reputation if it remains hidden. They may accordingly opt to either covertly bury such private transgressions or punish the transgressor in discreet ways that would keep the transgression hidden. Group members should be reluctant to publicly denounce private ingroup transgressions because doing so would publicize the ingroup transgression to outsiders and pose a threat to their group's reputation. In short, group members intending to advance their group's reputation should be motivated to publicly punish those who committed public transgressions and either protect or privately punish group members who committed private transgressions. To our knowledge, this is the first research examining the effect of publicness of ingroup transgressions on group members' responses to ingroup transgressions.

2. Amplifiers of reputation-protective instincts: identity fusion and situated responsibility

Who should be most prone to protecting the group's reputation in response to ingroup moral transgressions? We expect that individuals who feel a sense of responsibility for their group should be especially sensitive to reputational threats and should subsequently prioritize their group's reputation when responding to ingroup transgressions. Feelings of responsibility for the group can arise from dispositional or situational factors. One such dispositional factor is identity fusion (Swann Jr, Jetten, Gómez, Whitehouse, & Bastian, 2012).[1] For individuals whose identities are strongly fused to the group, their group identity is a core aspect of who they are (Swann Jr, Gómez, Seyle, Morales, & Huici, 2009) and is a chronically accessible determinant of their thoughts and emotions (Aquino et al., 2007). As such, individuals who are strongly fused with a group are particularly sensitive to threats faced by their group (Talaifar & Swann, 2019) and are apt to enact behaviors that advance their group's interests even when such behaviors are morally questionable (Fredman & Bastian, 2017) or involve personal sacrifices (Swann Jr et al., 2009). Ingroup moral violations deeply disturb strongly fused individuals (Buhrmester, 2013) and should motivate them to endorse responses that aid the group's reputation. That is, strongly fused group members should be most likely to publicly punish public transgressions and overlook or hide private transgressions by other group members.

Feelings of responsibility for one's group can also be situationally induced when individuals hold positions that allow them to control and take accountability for group outcomes (e.g., group leaders; Marques, Abrams, et al., 1998). Holding such positions would make group interests salient and would subsequently bias individuals toward actions that affirm group values and group reputation (Marques, Abrams, et al., 1998). Further, even if individuals holding positions of responsibility are unconcerned with the group per se, they may still work to protect the group because it in their own best interests as highly visible group members (Hornsey et al., 2005). Therefore, when individuals are situationally induced to feel responsible for their group, they should respond to transgressive group members in ways that maximize their group's reputation. In short, we expect that feelings of responsibility for the group, regardless of whether such feelings stem from one's chronic feelings of identity fusion or roles assigned to them in their immediate situational context, should induce people to respond to ingroup transgressions in ways that aid the group's reputation.

3. Overview of research

To address the foregoing ideas, we conducted four studies. In all studies, participants read hypothetical scenarios describing a moral violation (i.e., tax evasion) committed by members of their political party. We selected tax evasion because it is generally considered to be unethical (McGee, 2006), but it is not so morally abhorrent as to evoke uniformly extreme reactions. Further, because tax evasion does not usually cause direct harm to specific victims, we could construct a scenario in which the transgression was completely private. We selected political party as our focal group because reputation is especially critical to the survival and success of political parties. To maximize participant interest in the study materials, we conducted Studies 1–3 during a period of heightened political engagement: the six months preceding the mid-term elections of 2018, when threats to party reputation would have presumably been consequential for both parties. All our studies were restricted to participants who identified as supporters of either the Democratic or the Republican Party.

Study 1 employed a two-factor mixed design: 3 (Publicness of transgression) X 3 (Publicness of punishment). All participants were induced to feel responsible for their group, and we hypothesized that in such situations, they would respond to ingroup transgressors in ways that would aid their group's reputation. Specifically, we expected that participants would want their party to publicly denounce fellow group members who committed public transgressions but that they would opt for relatively discreet ways of punishment after private transgressions. Studies 2a and 2b, using a one-factor between-subjects design, examined participants' responses to private versus public ingroup transgressions in the absence of situationally-induced feelings of responsibility for the group. In such contexts, we expected individual differences in chronic feelings of responsibility for the group (conceptualized here as identity fusion) to predict reputation-protective responses to ingroup transgressions. Specifically, we expected that fused individuals would be most likely to endorse publicly punishing ingroup transgressors after public transgressions. We also hypothesized

[1] There may be other dispositional sources of responsibility for the group including group identification (Tajfel & Turner, 1979). We measured fusion because the verbal fusion measure (Gómez et al., 2011) has been especially potent in predicting efforts to protect the group through extreme pro-group behaviors.

that they would be especially likely to contemplate extreme actions aiming to hide private transgressions, such as destroying or tampering with incriminating evidence. In Studies 3 and 4, in addition to manipulating the publicness of ingroup transgressions, we also manipulated situated responsibility for the group, resulting in a two-factor between-subjects design: 2(Publicness of transgression) X 2(Situated responsibility). We hypothesized that individuals who were either experimentally induced to feel responsible for their group or who were strongly fused with their group would endorse publicly punishing group members who committed public, relative to private, transgressions. Study 4 also measured participants' reputation-protective motivation to determine whether individuals with a high motivation to protect group reputation were most likely to show the above-described effects of publicness of transgression.

Let us add a few methodological/statistical notes. For all of the studies, we report all measures, manipulations, and exclusions. Following the studies, we present a pooled analysis of all our samples. Finally, we present all study materials and report cell means and standard deviations, inter-variable correlations, and distributions of excluded participants across conditions in the Supplementary Online Materials (SOM).

4. Study 1

Study 1 examined whether group members occupying positions of responsibility are influenced by reputational concerns when deciding how to respond to the moral violations of group members. Participants imagined that they held a position of responsibility in a group and then read about a hypothetical scenario in which they encountered a moral transgression of an ingroup member which was known either (a) only to them (i.e., private transgression), or (b) to them and other group members (i.e., semi-private transgression), or (c) to members and nonmembers (i.e., public transgression). Participants subsequently specified how they would respond to the ingroup transgression. We hypothesized that participants would opt for publicly punishing the ingroup transgressor after public transgressions but would select relatively discreet punitive processes when responding to private transgressions. We also expected that participants would be reluctant to even discuss private ingroup transgressions with other group members if there was a risk of outsiders finding out about it (see SOM for related analyses). Finally, we tested whether strongly fused party members would be particularly likely to show these effects.

4.1. Methods

4.1.1. Participants

We recruited 444 participants from the United States using Amazon's Mechanical Turk (MTurk; minimum HIT approval rate = 95%). In this and all following studies, sample size was determined prior to data analysis. This study was run in July 2018 just before reports of corruption of MTurk's participant pool (TurkPrime, 2018) first surfaced in online researcher communities. As recommended by several researchers (Dennis, Goodson, & Pearson, 2018; TurkPrime, 2018), we excluded responses whose IP address were identified as being located outside the US ($N = 26$), whose qualitative responses (see "Other Measures" below) were flagged by two blind judges as unintelligible ($N = 28$; agreement rate = 94.13%), and who failed an attention-check question ($N = 43$; see SOM). In this and all following studies, we retained only the first response from each IP address to eliminate the possibility of single respondents completing the survey twice.

In all of our studies, participants were asked to select the political party that they identified with, and those who indicated that they did not support either the Democratic or Republican party were eliminated. The final sample had 347 participants ($\mathrm{N_{male}} = 136$; $\mathrm{N_{female}} = 153$; $\mathrm{N_{other}} = 1$; $\mathrm{N_{unknown}} = 57$; $M_{\mathrm{age}} = 37.90$; $SD_{\mathrm{age}} = 12.93$, 60.81% Democrat). A sensitivity analysis revealed that our sample had 80% power at $\alpha = 0.05$ to detect a 3 × 3 mixed effects interaction of a minimum size of $f = 0.10$ assuming the mean correlation among measured variables to be 0.15. Note that none of the effects reported in this or the following studies was moderated by political party membership.

4.1.2. Procedure

4.1.2.1. Identity fusion. Participants completed the verbal fusion scale (Gómez et al., 2011) measuring identity fusion with political party (e.g., "I am one with the Democratic/Republican party"). They rated items on a 7-point scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). Responses to all items were averaged ($M = 4.12$, $SD = 1.49$; $\alpha = 0.93$, 95% CI [0.92, 0.94]).

4.1.2.2. Between-subjects manipulation of publicness of transgression. Participants were asked to imagine they had been selected by their political party to participate in local-level party meetings and make decisions on behalf of the party. They learned further than they had encountered evidence implicating a politician of their political party in tax fraud. We manipulated the publicness of the evidence. In the **private transgression** condition ($N = 114$), participants learned that they were the only one privy to information regarding the party member's tax fraud. In the **semi-private transgression** condition ($N = 119$), the participant and some core members of their political party knew about the transgression. In the **public transgression** condition (N = 114), information regarding the party member's tax fraud was already publicized widely by the media.

4.1.2.3. Punishing the transgressive group member. Participants answered a series of questions indicating their willingness to endorse punishing the transgressive politician using methods of varying degrees of publicness on 7-point scales ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). Participants indicated their likelihood of urging their party to **privately punish** the transgressor (e.g., "I would condemn the party member's behavior in a private and confidential conversation with him"; $M = 4.80$, $SD = 1.69$; $\alpha = 0.86$, 95% CI [0.83, 0.89]). They also rated their willingness to urge their party to take action in a manner known only to the party's decision-makers (e.g., "I would advise my party to punish the member via a confidential process known only to core party members"), which was our index of **semi-private punishment** ($M = 4.20$, $SD = 1.77$; $\alpha = 0.86$, 95% CI [0.83, 0.89]). Finally, they rated how likely they were to urge their party to **publicly punish** the party member transgressor (e.g., "I would recommend that my party go public by reporting the party member to federal agents"; $M = 4.49$, $SD = 1.80$; $\alpha = 0.89$, 95% CI [0.87, 0.91]). We randomized the order in which participants answered these three sets of questions.

4.1.2.4. Other measures. Finally, participants completed an open-ended response revealing their thoughts on the incident. They then provided demographic information and were debriefed.

4.2. Results

4.2.1. Punishing the transgressive group member

We wanted to determine whether participants, who were all assigned to positions of responsibility, considered consequences for group reputation while selecting punitive processes against transgressive group members. Our hypothesis was that participants would aid group reputation by selecting public punishment after public transgressions and relatively private punitive processes after private transgressions. To test this, we examined a mixed effects model predicting willingness to endorse punishment, with publicness of transgression (private vs semi-private vs public) as a between-subjects factor and publicness of punishment (private vs semi-private vs public) as a within-subjects factor. As hypothesized, the 2-way interaction of publicness of transgression and publicness of punishment was significant ($F(4, 684) = 4$, $p = .004$).

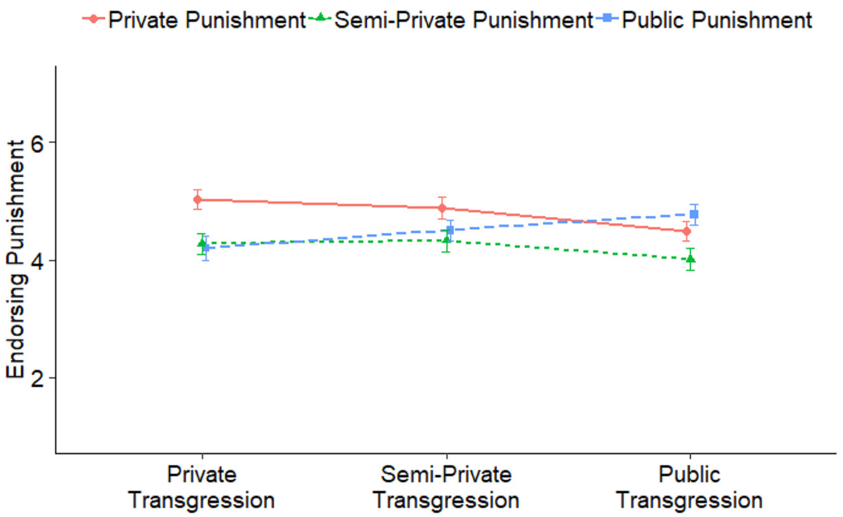

Fig. 1. Endorsing punishment as a function of publicness of ingroup transgression and publicness of punishment. Error bars indicate 95% confidence intervals.

As evident from Fig. 1, when participants encountered a public transgression, they preferred public punishment ($M = 4.77$, $SD = 1.73$) to semi-private punishment ($M = 4.01$, $SD = 1.83$, $t(684) = -3.58$, $p < .001$, $d_o = 0.47$).[2] Further, as indicated by the blue line, public punishment was more likely after public, relative to private, transgressions ($t(344) = 2.48$, $p = .01$, $d_o = 0.36$). After a private transgression, participants preferred private punishment ($M = 5.03$, $SD = 1.64$) to semi-private ($M = 4.27$, $SD = 1.72$, $t(684) = -3.54$, $p < .001$, $d_o = 0.47$) or public punishment ($M = 4.20$, $SD = 1.92$, $t(684) = -3.89$, $p < .001$, $d_o = 0.52$) In sum, participants seemed to generally prefer public punishment after public transgressions and private punishment after private transgressions.

4.2.2. *Effects of identity fusion*

Noting that the above reported effects were strongest when comparing private and public transgressions, we conducted an exploratory analysis based on excluding the semi-private transgression condition. In a mixed-effects model with publicness of transgression (private vs. public), fusion with party, and publicness of punishment (private vs semi-private vs public) as predictors, the 3-way interaction term was marginally significant ($b = 0.35$, $t(445) = 1.80$, $p = .07$; see Fig. 2). We conducted simple effects tests to determine if the pattern of the marginally significant interaction confirmed prediction.

As revealed in the third panel of Fig. 2, strongly fused people (i.e., at +1 SD fusion) were more likely to endorse public punishment after a public transgression than a private one ($b = 1.02$, $t(224) = 3.30$, $p = .001$). Whereas, on encountering private transgressions, they preferred private over public punishment ($b = -1.36$, $t(445) = -4.81$, $p < .001$). However, the willingness of weakly fused participants (i.e., at −1 SD fusion) to endorse private, semi-private, or public punishment did not depend on the publicness of their group member's transgression ($0.13 > ps < 0.73$). While these analyses are underpowered, the findings provide preliminary evidence that strongly fused people are especially likely to endorse reputation-protective responses to ingroup transgressions.

4.3. *Discussion*

In Study 1, individuals occupying a position of responsibility for their group responded to ingroup transgressions in ways that would aid their group's reputation. After public transgressions, they preferred public punishment over private punishment; after private transgressions, they preferred private punishment over public punishment. Our data also provides some preliminary evidence that strongly fused individuals were particularly likely to show the above-described effects of publicness. The fusion effects were relatively weak, possibly because all participants in this study were encouraged to feel responsible for the group, thereby rendering it difficult to observe any additional effects of fusion. Moreover, our sample did not have sufficient power to detect a 3-way interaction. We addressed these issues in the following studies.

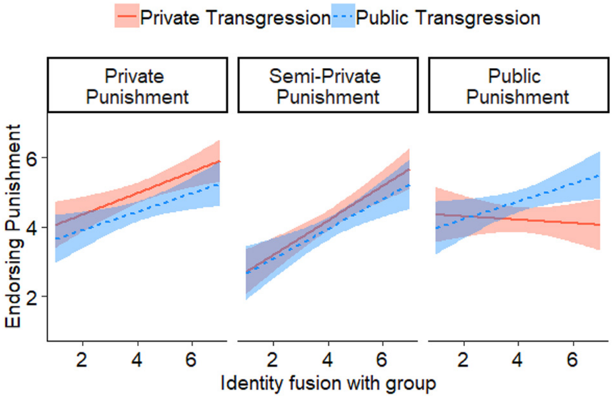

Fig. 2. Endorsing punishment as a function of fusion, publicness of transgression and publicness of context. Bands indicate 95% confidence intervals.

5. Studies 2a and 2b

We probed the fusion-related effects found in Study 1 by testing a pre-registered set of hypotheses (https://osf.io/eaj7f/?view_only=1cdb34ead54f480b88ef7ff60706441f) in two samples. To increase the study's power to detect fusion-related effects, we simplified the design in two ways. First, we limited our manipulation of publicness of transgression to two levels: private versus public transgressions. Second, instead of treating publicness of a punishment as a within-subject factor, we measured public punishment alone. Further, we did not assign positions of responsibility to participants in this study.

We tested three pre-registered hypotheses. First, we expected that strongly fused individuals would be especially likely to urge their group to publicly punish ingroup transgressors after public, relative to private, transgressions. We expected weakly fused individuals to either not show this pattern or show the opposite pattern. Second, we considered

[2] d_o refers to operative effect size (Judd, Westfall, & Kenny, 2017) computed as the difference between conditions divided by the square root of the estimated residual error

the possibility that strongly fused individuals might also be more willing to directly report ingroup transgressors to federal agents after public, as opposed to private, transgressions. If strongly fused individuals want their group to publicly punish public transgressors but are reluctant to do so themselves, this would support our claim that concerns for the group's moral reputation underlie strongly fused members' responses to ingroup transgressions.

Third, we expected that strongly fused people, on encountering a private transgression by a fellow group member, would be especially willing to endorse actions designed to hide the transgressions in an effort to protect their group's reputation. To test this hypothesis, we measured people's willingness to destroy (Study 2a) and tamper with (Study 2b) evidence incriminating a member of their group.

5.1. Methods

5.1.1. Participants

We recruited 538 participants in Study 2a and 529 participants in Study 2b using MTurk. As reported in our pre-registration, we based our sample sizes on a power analysis revealing that 518 participants were needed to detect the interaction effect of fusion and publicness reported in Study 1 ($f = 0.12$) with 80% power. We restricted our studies to participants in the US with a respectable HIT approval rate (99% in Study 2a; 98% in Study 2b) and history (> 500 approved HITs). In addition, to deter non-US MTurkers from participating via server farms, Study 2b used a novel mechanism offered by TurkPrime to block participants from suspicious geolocations (Robinson, 2018) and excluded two participants whose IP addresses were mapped to locations outside the US. We did not identify any such participants from Study 2a. Further, following the pre-registered exclusion criteria, we excluded participants if they failed either of two attention checks (see SOM) embedded in the surveys ($N = 14$ in Study 2a; $N = 73$ in Study 2b) or if their response to an open-ended prompt (see "Other Measures" below) was judged by two blind raters as unintelligible ($N = 4$ in Study 2a; $N = 8$ in Study 2b; agreement rate = 99.62% in Study 2a and 98.24% in Study 2b). The final samples had 520 participants ($N_{male} = 196$; $N_{female} = 278$; $N_{other} = 1$; $N_{unknown} = 45$; $M_{age} = 38.70$; $SD_{age} = 12.48$, 64.42% Democrat) in Study 2a and 446 participants ($N_{male} = 161$; $N_{female} = 251$; $N_{other} = 2$; $N_{unknown} = 32$; $M_{age} = 36.60$; $SD_{age} = 11.02$, 60.09% Democrat) in Study 2b. Sensitivity analysis revealed that our final samples had 80% power at $\alpha = 0.05$ to detect interaction effects of size $f = 0.12$ in Study 2a and $f = 0.13$ in Study 2b.

5.1.2. Procedure

5.1.2.1. Identity fusion. Participants completed the verbal fusion scale measuring fusion with political party (Gómez et al., 2011) in both studies 2a ($M = 3.92$, $SD = 1.45$; $\alpha = 0.93$, 95% CI [0.92, 0.94]) and 2b ($M = 4.04$, $SD = 1.39$; $\alpha = 0.92$, 95% CI [0.91, 0.94]).

5.1.2.2. Manipulation of publicness of transgression. Participants then read a vignette describing a hypothetical scenario in which the participant, while browsing through an online archive, encounters documents implicating a politician of their political party in tax fraud. Unlike in Study 1, the vignette did not assign a position of responsibility to the participant. Participants in each condition received different information about how many people knew about the tax fraud. In the **private transgression** condition (Study 2a: $N = 265$; Study 2b: $N = 224$), the vignette said that the participant was the only one privy to information regarding the party member's tax fraud. In the **public transgression** condition (Study 2a: $N = 255$; Study 2b: $N = 222$), information regarding the tax fraud was already publicized widely by the media. There were minor differences between vignettes used in Study 2a and Study 2b (see SOM).

5.1.2.3. Urging the group to publicly punish the transgressive group member. Participants answered two questions indicating their likelihood of urging their party to publicly punish the transgressive party member (e.g., "I would write to my local party office urging that they go public by reporting the party member to federal agents"). Participants rated these items on a 7-point scale (Study 2a: $M = 4.04$, $SD = 1.87$; $\alpha = 0.81$, 95% CI [0.78, 0.84]; Study 2b: $M = 4.08$, $SD = 1.85$; $\alpha = 0.83$, 95% CI [0.80, 0.86]).

5.1.2.4. Directly reporting the transgressive group member. Participants also rated two items on their willingness to directly report the transgressor to federal agents unmediated by their group (e.g., "I would report the party member to federal agents by sharing all the evidence I found") on a 7-point scale (Study 2a: $M = 4.35$, $SD = 2.02$; $\alpha = 0.95$, 95% CI [0.94, 0.96]; Study 2b: $M = 4.32$, $SD = 1.97$; $\alpha = 0.96$, 95% CI [0.96, 0.97]).

5.1.2.5. Hiding evidence incriminating the transgressive group member. In Study 2a, participants rated their willingness to destroy evidence incriminating an ingroup transgressor (e.g., "I would delete the documents from the online archive") on a seven-point scale ($M = 1.64$, $SD = 1.26$; $\alpha = 0.96$, 95% CI [0.96, 0.97]). This measure suffered from a floor effect and a heavy positive skew, with 68.85% of the sample at the lowest possible score. To evade this problem in Study 2b, we opted for a milder measure: Willingness to tamper with evidence (e.g., "I would try to move the document to an offline server"). This measure too was heavily skewed ($M = 2.29$, $SD = 1.64$; $\alpha = 0.90$, 95% CI [0.88, 0.92]), with 49.33% of the sample at the lowest score.

In both studies, we randomized the order in which participants answered the questions regarding urging the group to publicly punish the transgressor and tampering with evidence.

5.1.2.6. Other measures. Finally, participants completed an open-ended response recording their thoughts on the incident before providing demographic information and being debriefed.

5.2. Results

5.2.1. Urging the group to publicly punish the transgressive group member

5.2.1.1. Main effect of publicness of transgression. In the absence of situational sources of responsibility, we expected only strongly fused group members ingroup transgressions to react in ways that maximized group reputation. This is why our pre-registration did not hypothesize a main effect of publicness of transgression on publicly punishing ingroup transgressors. In Study 2a, we found a marginal effect of publicness of transgression ($p = .09$) in the opposite direction of the effect in Study 1, and in Study 2b, we found no effect ($p = .97$). In the absence of a situational source of responsibility for the group, members are, in general, not swayed by reputational consequences for their group.

5.2.1.2. Interaction of fusion and publicness of transgression. We expected strongly fused people to be most likely to endorse the reputation-protective response of wanting the group to publicly punish transgressors after public transgressions.

In Study 2a, we found a significant interaction of fusion and publicness of transgression ($\beta = 0.13$, $t(515) = 2.22$, $p = .03$, $f = 0.10$). Simple effects analyses indicated that weakly fused individuals (−1 SD) were more likely to endorse public punishment after private transgressions than public transgressions ($\beta = -0.34$, $t(515) = -2.76$, $p = .006$), which is a response that would harm group reputation. Strongly fused (+1 SD) individuals did not show this effect ($p = .70$). Further, fusion was associated public punishment after a public transgression ($\beta = 0.21$, $t(515) = 3.39$, $p < .001$) but not a private one ($p = .74$), highlighting strongly fused individuals' preference for reputation-friendly responses.

In Study 2b, a marginally significant interaction effect of fusion and publicness of transgression emerged, ($\beta = 0.12$, $t(441) = 1.85$, $p = .07$, $f = 0.09$). Simple effects analyses paralleled our findings from Study 2a.

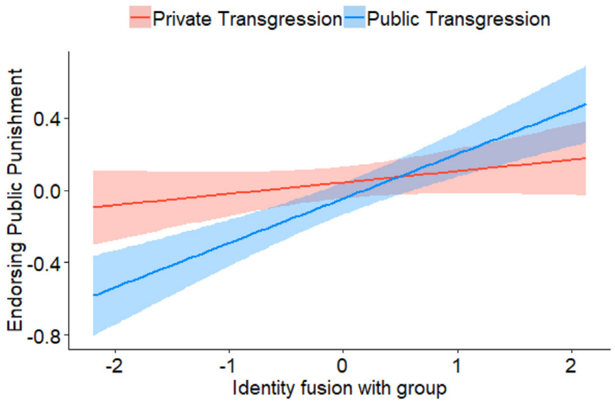

Fig. 3. Endorsing public punishment as a function of fusion and publicness of transgression in a combined sample from Studies 2a and 2b. Bands indicate 95% confidence intervals.

Strongly fused (+1 SD; β = 0.15, t(441) = 1.11, p = .27) individuals showed a non-significant preference to endorse public punishment after public, relative to private, transgressions. Weakly fused individuals showed the opposite albeit non-significant effect (−1 SD; β = −0.20, t(441) = −1.51, p = .13). As in Study 2a, fusion was associated with urging one's group to publicly punish an ingroup transgressor after public transgressions (β = 0.28; t(441) = 4.07, p < .001) but only marginally associated after private transgressions (β = 0.11; t(441) = 1.71, p = .09).

Pooling the two samples to increase power revealed a significant interaction of fusion and publicness of transgression (β = 0.13, t(959) = 2.86, p = .004, f = 0.10; see Fig. 3). Using Fischer's method of aggregating p-values from Studies 2a and 2b, we found the joint probability of the interaction of fusion and publicness to be p = .01.

5.2.2. *Directly reporting the transgressive group member*

We did not find a significant interaction of fusion and publicness on directly reporting transgressive group members in either Study 2a (p = .24) or Study 2b (p = .27), suggesting that when strongly fused members encounter public ingroup transgressions, they are especially motivated to have their group publicly distance itself from the transgressor but not do so themselves. This finding strengthens our argument that public punishment enacted by the group serves a special function of maximizing group reputation.

5.2.3. *Hiding evidence incriminating the transgressive group member*

Scores on our measures of destroying (Study 2a) and tampering with (Study 2b) evidence severely violated normality. We addressed this problem by testing our hypothesis using an ordinal logistic regression.

In Study 2a, an ordinal regression revealed that the interaction of fusion and publicness had no effect on willingness to destroy evidence (p = .28). However, fusion was significantly associated with destroying evidence (b = 0.42, t(517) =4.28, OR = 1.52, 95% CI = [1.26,1.84], p < .001) indicating that fused individuals were most willing to destroy evidence.

An ordinal logistic regression in Study 2b similarly revealed no interaction effect of fusion and publicness of transgression (p = .17). Further, confirming our finding in Study 2a, fused members were most willing to endorse tampering with evidence (b = 0.29, t(443) =3.12, OR = 1.34, 95% CI = [1.12, 1.63], p = .002).

5.3. *Discussion*

Studies 2a and 2b revealed that among people who don't occupy positions of responsibility, strongly fused group members were more likely than others to respond in ways that aid group reputation. That is, in both studies, fusion was associated with urging one's group to publicly punish ingroup transgressors after public, more than private, transgressions. This is presumably because having one's group publicly punish a transgressive member would aid group reputation after public transgressions but threaten group reputation after private ones. On the contrary, weakly fused people opted for responses that would threaten, rather than protect, group reputation. They were more likely to urge their group to publicly punish transgressors after private transgressions, which may have been either because these individuals thought of it as unnecessary to report transgressions that were already public or because they felt like the onus was on them to take moral action as they were the only bystander (Darley & Latané, 1968) in the case of private transgressions. Further, while strongly fused individuals were most likely to urge their group to publicly punish group members who committed public transgressions, they were not similarly willing to directly report ingroup transgressors to federal agents in such situations. That is, fused individuals want their party to symbolically distance the transgressor rather than do so themselves, which supports our claim that concerns for group reputation underlie their responses to ingroup transgressions.

Further, strongly fused people were most likely to endorse extreme behaviors motivated to protect group reputation such as destroying or tampering with evidence. Inconsistent with our expectations, we did not find evidence that this effect was moderated by the publicness of ingroup transgressions. In Study 3, we attempted to replicate these findings and also experimentally examine the role of situated responsibility in responses to ingroup moral transgressions.

6. Study 3

Participants in Study 1, who were all situationally induced to feel responsible for their group, responded to ingroup moral transgressions in ways that aided their group's reputation. Conversely, participants in Study 2, none of whom were exposed to situational sources of responsibility, did not generally opt for reputation-protective responses. Only strongly fused individuals in Study 2 responded to transgressions in reputation-protective ways. In Study 3, we sought to systematically test our hypothesis that situational feelings of responsibility arising from occupying positions of responsibility within the group impacts participants' responses to ingroup transgressions. We also wanted to replicate the fusion effects uncovered in the preceding studies. We focused this study on the two responses that were especially relevant to our hypothesized mechanism regarding group reputation: (a) urging one's group to publicly punish ingroup transgressors, and (b) tampering with evidence incriminating transgressors.

6.1. *Methods*

6.1.1. *Participants*

We recruited a sample of 979 MTurkers from the US with a respectable HIT approval rating (> 98%) and history (> 500 accepted HITs). We used the mechanism described in Study 2b to block participants from suspicious geolocations and excluded two participants whose IP addresses mapped to locations outside the US. Further, participants who either failed an attention-check (N = 17) or whose qualitative responses were evaluated by two blind judges (agreement rate = 97.90%) as unintelligible (N = 24) were excluded. We were finally left with a sample of 936 participants (N_{male} = 353; N_{female} = 578; N_{other} = 4; $N_{unknown}$ = 1; M_{age} = 36.50; SD_{age} = 12.01, 63.25% Democrat). A sensitivity analysis revealed that our sample had 80% power to detect interaction effects of a minimum size f = 0.09, which is approximately the size of the effects reported in preceding studies (0.09 < = f < = 0.12).

6.1.2. *Procedure*

6.1.2.1. Identity fusion. Participants first completed the verbal fusion scale measuring fusion with political party (Gómez et al., 2011;

$M = 3.89$, $SD = 1.34$; α = 0.92, 95% CI [0.91, 0.92]).

6.1.2.2. Design. The study employed a 2 (Situated Responsibility: responsibility vs control) X 2 (Publicness of Transgression: private vs public) between-subjects design. Participants read a vignette describing a hypothetical scenario in which they happened to encounter evidence implicating a politician of their political party in tax fraud.

In the **situated responsibility** condition ($N = 452$), the vignette described a scenario in which the participant had been picked by their political party to attend local-level party meetings for a month. Participants were also told that they would be authorized to make decisions on behalf of the party during those meetings. In the **control** condition ($N = 484$), the vignette made no mention of this.

Similar to our previous studies, the publicness of transgression manipulation varied the information participants read about the number of people who knew about the transgression. In the **private transgression** condition ($N = 334$), the vignette added that only the participant knew of the party member's tax fraud. In the **public transgression** condition ($N = 602$), they read that the party member's tax fraud was already widely publicized. We recruited more participants for the public condition because we wanted to test an additional exploratory hypothesis regarding the effect of situated responsibility on strongly fused individuals' responses to public transgressions (refer to SOM).

6.1.2.3. Urging the group to publicly punish the transgressive group member. Participants answered two questions indicating their likelihood of urging their party to publicly punish the transgressing politician (e.g., "I would urge my party to go public by reporting the party member to federal agents"). Participants rated these items on a 7-point scale ($M = 5.05$, $SD = 1.61$; α = 0.67, 95% CI [0.62, 0.71]).

6.1.2.4. Urging the group to hide evidence incriminating the transgressive group member. Participants then rated two items measuring their willingness to urge their party to tamper with evidence they came across (e.g., "I would urge my party to try to move the documents to an offline server"; $M = 3.13$, $SD = 1.76$; α = 0.90, 95% CI [0.89, 0.91]). Unlike in Studies 2a and 2b, this measure did not show a floor effect and was only mildly skewed (24.68% of the sample at the lowest score).

Finally, participants answered an open-ended question, provided demographic information and were debriefed.

6.2. Results

6.2.1. Urging the group to publicly punish the transgressive group member

6.2.1.1. Interaction of fusion or situated responsibility with publicness of transgression. To examine whether individuals who were situationally induced to feel responsible for their group or strongly fused with the group were especially likely to endorse publicly punishing the transgressor after public, as opposed to private, transgressions, we examined a model including both the hypothesized interaction terms (i.e., (a) situated responsibility and publicness, and (b) fusion and publicness). We found a marginal interaction effect of situated responsibility and publicness of transgression (β = 0.12, $t(930) = 1.84$, $p = .07$, $f = 0.06$) and a significant interaction effect of fusion and publicness of transgression (β = 0.11, $t(930) = 1.93$, $p = .05$, $f = 0.06$).

Breaking down these interactions first revealed that people who were experimentally induced to feel responsible for their group were more likely to urge their party to publicly punish a transgressor after public transgressions ($M = 5.27$, $SD = 1.42$), as opposed to private transgressions ($M = 4.70$, $SD = 1.60$, $F(1, 450) = 15.60$, $p < .001$), while those in the control condition did not show this effect ($p = .24$; See Fig. 4a).

Further, as indicated in Fig. 4b, strongly fused (+1 SD; β = 0.37, $t(932) = 3.75$, $p < .001$) individuals were more likely to endorse publicly punishing the transgressor after public transgressions relative to private ones. Weakly fused individuals showed no such difference (−1 SD; $p = .26$).

6.2.2. Urging the group to hide evidence incriminating the transgressive group member

A significant main effect of fusion (β = 0.16, $t(931) = 5.08$, $p < .001, f = 0.17$) indicated that strongly fused individuals were most likely to tamper with evidence incriminating ingroup transgressors. Neither fusion nor situated responsibility interacted with publicness ($ps > 0.72$) and the main effect of the manipulation of situated responsibility was not significant ($p = .38$).

6.3. Discussion

Findings from Study 3 replicated the fusion effects found in Studies 1, 2a and 2b. In addition, Study 3 confirmed our speculation that people who are situationally induced to feel responsible for their group respond to ingroup transgressions in a manner that resembles strongly fused group members. While the fusion effects are generally consistent across studies, there are minor differences in the nature of the interaction. For example, strongly fused individuals showed an effect of publicness in Studies 1 and 3 (they preferred public punishment after public, more than private, transgressions) but it was weakly fused individuals who showed an effect of publicness in Studies 2a and 2b albeit in the opposite direction (they preferred public punishment after private, more than public, transgressions). We believe that these differences are due to differences in study design and outcome measure. In Studies 1 and 3, in which at least half the sample was assigned to positions of responsibility and measures of public punishment involved relatively abstract actions ("urging the party"), participants may have felt more inclined to endorse public punishment (notice the higher overall means in Study 1 and 3). Given that public punishment was relatively easy to endorse in these studies, inaction after a private transgression (presumably to protect reputation) was a difficult choice that only fused individuals preferred. In contrast, in Studies 2a and 2b, in which participants were not assigned positions of responsibility and the measure of public punishment involved a concrete, effortful action ("writing to the party"), participants may have been reluctant to endorse public punishment, and especially so among weakly fused people in the public transgression condition because they may have felt that there is little reason to report a transgression that others already know about. While nature of the interaction may have depended on the vignette and measures of the outcome variable, the key finding is that strongly fused individuals consistently showed a higher public-private difference than weakly fused individuals in their willingness to endorse public punishment.

7. Study 4

The first three studies provide converging evidence for our idea that individuals who are either fused with their group or who occupy positions of responsibility within the group respond to ingroup transgressions in ways that aid their group's reputation. Study 4 aimed to test whether these effects are indeed driven by individuals' motivation to protect group reputation. As proposed in the model displayed in Fig. 5, we expected that individuals who were either experimentally induced to feel responsible for the group or fused with the group would be motivated to protect group reputation. Such a motivation to protect group reputation would be associated with wanting the group to publicly punish ingroup transgressors more after public transgressions relative to private ones.

The study also attempted to test two alternate hypotheses. First, participants may have perceived public transgressions to be especially immoral, and such perceptions of immorality may have induced them to endorse public punishment after public transgressions. Second, it is

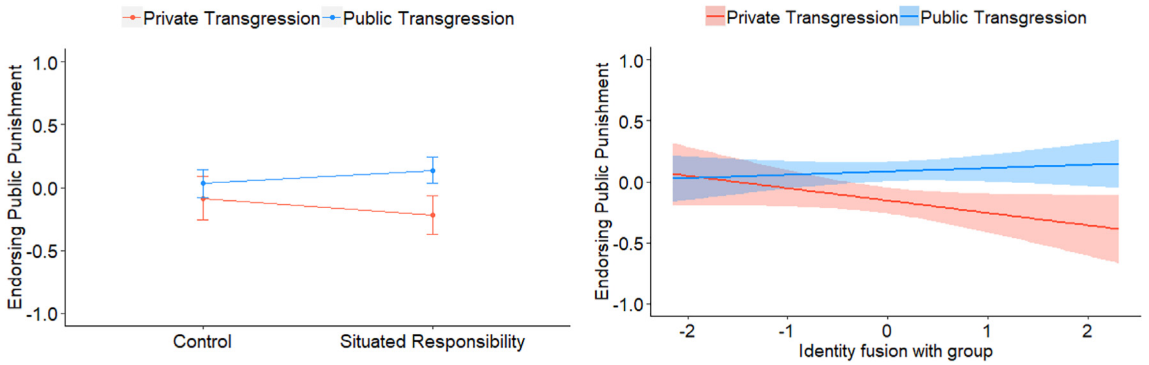

Fig. 4. Endorsing public punishment as a function of the interaction of situated responsibility (Fig. 4a) or fusion (Fig. 4b) with publicness of transgression in Study 3. Bands and error bars indicate 95% confidence intervals.

possible that the manipulation of publicness inadvertently altered the activated group context. Perhaps public transgressions activated an intergroup context by making outgroups salient and private transgressions activated an intragroup context. This difference in activated group context may have produced the effects of publicness.

7.1. Methods

7.1.1. Participants

We recruited a sample of 685 participants from the United States using Prolific Academic (approval rate over 95%). Participants who failed either of two attention check questions were excluded ($N = 7$). The final sample had 678 participants ($N_{male} = 290$; $N_{female} = 364$; $N_{other} = 14$; $N_{unknown} = 10$; $M_{age} = 33.7$; $SD_{age} = 12.42$, 78.2% Democrat), which had 80% power to detect interaction effects of a minimum size $f = 0.11$.

7.1.2. Procedure

7.1.2.1. Identity fusion. Participants first completed the verbal fusion scale measuring fusion with political party (Gómez et al., 2011; $M = 3.84$, $SD = 1.36$, α = 0.92, 95% CI [0.91, 0.93]).

7.1.2.2. Design. As in Study 3, a 2 (Situated Responsibility: responsibility vs control) X 2 (Publicness of Transgression: private vs public) between-subjects design was used. Participants read a vignette describing the same ingroup transgression as in the previous studies.

In the **situated responsibility** condition ($N = 348$), but not in the **control** condition ($N = 330$), participants learned that they were in a position of authority and could make decisions on behalf of the party. A manipulation check question (see SOM) revealed that participants did indeed perceive higher responsibility in the experimental condition ($M = 5.70$, $SD = 1.49$) than in the control condition ($M = 4.61$, $SD = 1.93$), $t(676) = 8.28$, $p < .001$, $d = 0.63$.

As in the previous studies, in the **private transgression** condition ($N = 336$), only the participant knew of the party member's tax fraud. In the **public transgression** condition ($N = 342$), the party member's tax fraud was already publicly known. A second manipulation check verified that participants perceived the transgression in the public transgression condition ($M = 5.77$, $SD = 1.43$) to be more public than in the private transgression condition ($M = 1.76$, $SD = 1.43$)), $t(676) = 36.50$, $p < .001$, $d = 2.80$.

7.1.2.3. Motivation to protect group reputation. Participants rated two items indicating how motivated they were to protect their group's reputation (e.g., "In this situation, I would want my party to make a decision that would protect its image") on a seven-point scale ($M = 4.07$, $SD = 1.85$; α = 0.85, 95% CI [0.82, 0.87]).

7.1.2.4. Perceptions of immorality. Participants also rated how immoral they perceived the transgression to be (e.g., "The party member's actions are immoral) on a seven-point scale ($M = 6.12$, $SD = 1.27$; α = 0.93, 95% CI [0.92, 0.94]).

7.1.2.5. Urging the group to publicly punish the transgressive group member. Participants rated the same items used in Study 3 measuring public punishment ($M = 5.64$, $SD = 1.48$; α = 0.80, 95% CI [0.77, 0.83]).

7.1.2.6. Urging the group to hide evidence incriminating the transgressive group member. Participants also rated the items used in Study 3 measuring their willingness to endorse hiding evidence ($M = 3.07$, $SD = 1.75$; α = 0.89, 95% CI [0.87, 0.90]).

7.1.2.7. Activated group context. Participants then rated the extent to which they considered the opinions of four groups – their own party's supporters, the opponent party's supporters, the electorate, and the media – when they contemplated their response to their party member's transgression. Each item was rated on a five-point scale (1 – *Not at all;* 5 – *A great deal).*

Finally, participants provided demographic information and were debriefed.

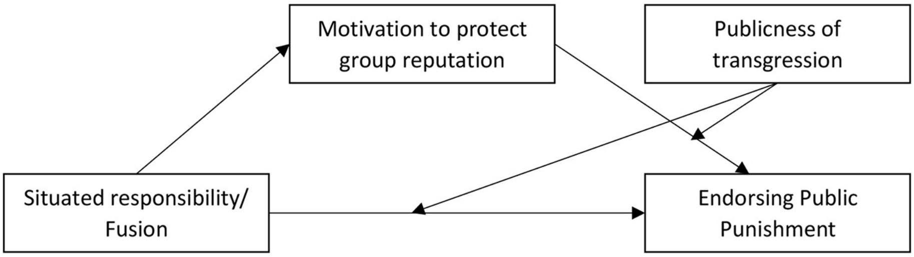
Motivation to protect group reputation
Publicness of transgression
Situated responsibility/ Fusion
Endorsing Public Punishment

Fig. 5. Theoretical moderated mediation model tested in Study 4.

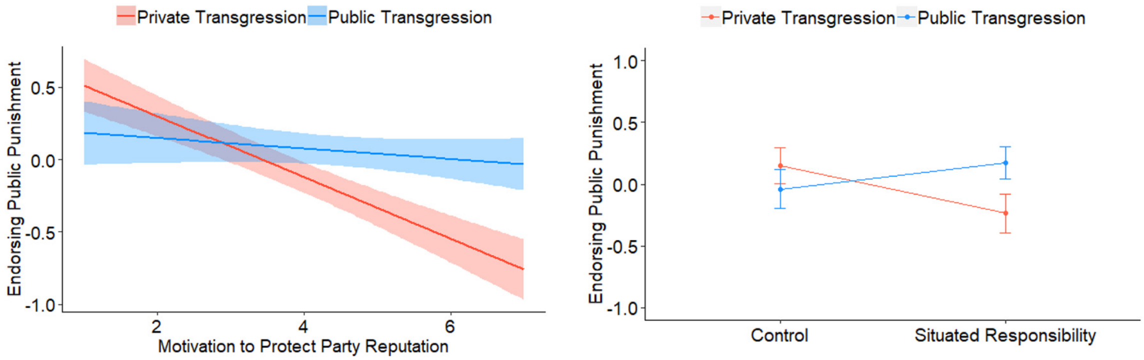

Fig. 6. Endorsing public punishment as a function of the interaction of reputation motivation (Fig. 6a) or situated responsibility (Fig. 6b) with publicness of transgression in Study 4. Bands and error bars indicate 95% confidence intervals.

7.2. Results

7.2.1. Urging the group to publicly punish the transgressive group member

We first tested the reputational hypothesis by examining the interaction of reputation-protective motivation and publicness. We then tested the moderated mediation model depicted in Fig. 5, first with situated responsibility as the predictor and then with fusion as the predictor.

7.2.1.1. Effects of reputation-protective motivation. We sought to determine whether individuals who reported high levels of reputation-protective motivation were more likely to endorse public punishment after public transgressions than private ones. Consistent with our hypothesis, the interaction of reputation motivation and publicness of transgression was significant (β = 0.23, *t*(674) = 4.35, *p* < .001, *f* = 0.17; see Fig. 6a) such that people who were highly motivated to protect group reputation (at +1 SD) were especially likely to show a public-private difference in their likelihood of endorsing public punishment, β = 0.54, *t*(674) = 5.07, *p* < .001. Those who were not motivated to protect group reputation (at −1 SD) did not show an effect of publicness of transgression (*p* = .28).

7.2.1.2. Effects of situated responsibility. We then examined the effects of the manipulation of situated responsibility. Replicating the finding from Study 3, we found a significant interaction effect of situated responsibility and publicness of transgression (*F*(1, 674) = 15.08, *p* < .001, *f* = 0.15; *c* * mod path in Fig. 5). As indicated in Fig. 6b, people who were experimentally induced to feel responsible for their group were more likely to urge their party to publicly punish a transgressor after public transgressions (*M* = 5.89, *SD* = 1.26) than private transgressions (*M* = 5.29, *SD* = 1.61), *F*(1, 346) = 14.75, *p* < .001. Participants in the control condition did not show this effect (*p* = .09). If anything, they showed the opposite effect (see Fig. 6b).

We tested the moderated mediation model depicted in Fig. 5 with situated responsibility as predictor. Guided by recent recommendations, we report each individual parameter in the model (Yzerbyt, Muller, Batailler, & Judd, 2018). First, testing the *a* path in the model revealed a significant effect of the situated responsibility manipulation on reputation motivation, *b* = 0.45, *t*(674) = 6.05, *p* < .001, *f* = 0.19, such that situationally responsible participants were especially likely to be motivated to protect group reputation. Second, we tested a model with the two moderated paths depicted in Fig. 5. The interaction of reputation-protective motivation and publicness (*b * mod* path) was significant (*b* = 0.28, *t*(672) = 3.60, *p* < .001), and so was interaction of situated responsibility and publicness (*c' * mod*; *b* = 0.44, *t*(672) = 2.92, *p* = .004). The indirect moderated mediation effect computed using 5000 Monte Carlo iterations was significant (IE = 0.12, 95% CI = [0.05, 0.21]), which provides evidence for the reputation hypothesis made in this paper.

7.2.1.3. Effects of fusion. We then examined whether individuals who were fused with their group were more likely than others to want to publicly punish the transgressor after public, as opposed to private, transgressions. Surprisingly, we did not find a significant interaction of fusion and publicness of transgression (*p* = .43; *c * mod* path). Even though the fusion effect was not significant, guided by assertions that there may be indirect effects even in the absence of a significant total effect (Zhao, Lynch Jr, & Chen, 2010), we tested the component paths in the proposed model (see Fig. 5). As hypothesized, participants who were strongly fused with their party reported higher levels of motivation to protect group reputation (*b* = 0.37, *t*(674) = 10.67, *p* < .001, *f* = 0.41; *a* path). Further, in a model with the two moderated paths (fusion * publicness; reputation-protective motivation * publicness), the *b * mod* path corresponding to the interaction of reputation-protective motivation and publicness was significant (*b* = 0.35, *t*(672) = 4.33, *p* < .001) such that people who were highly motivated to protect group reputation were especially likely to show a public-private difference in their likelihood of endorsing public punishment. Given that both the *a* path and the *b * mod* path were significant, the indirect effect depicted in Fig. 5 was significant (IE = 0.13, 95% CI = [0.07, 0.20]) even though the direct effect (*c * mod* path) was not.

7.2.1.4. Testing alternative explanations

7.2.1.4.1. Perceptions of immorality. To test whether people perceived public transgressions to be more immoral than private transgressions, we conducted a *t*-test and found no effect (*p* = .73). Participants' perceptions of immorality did not depend on how public the transgression was, which suggests that the reported effects of publicness were not driven by differences in perceived immorality.

7.2.1.4.2. Activated group context. Next we analyzed participants' ratings regarding the extent to which they were concerned about the opinions of four different groups – their own party's supporters, the opponent party's supporters, the electorate, and the media. We conducted a mixed effects model with publicness as between-subjects predictor and the group as within-subjects factor. The interaction of activated group and publicness was not significant (*p* = .92), indicating that publicness of the transgression did not influence which group participants were concerned about. This finding reduces concerns that differences in activated group context underlie the reported effects of publicness. Interestingly, we found a main effect of activated group context, *F*(3, 2016) = 84.23, *p* < .001, *f* = 0.35, indicating that participants cared most about the media (*M* = 3.16, *SD* = 1.40), followed by the electorate (*M* = 2.97, *SD* = 1.35), their own party's supporters (*M* = 2.84, *SD* = 1.33), and finally, the opposite party's supporters (*M* = 2.38, *SD* = 1.36).

7.2.2. Urging the group to hide evidence incriminating the transgressive group member

As in Studies 2–3, strongly fused individuals were most likely to endorse hiding evidence incriminating ingroup transgressors (β = 0.18, $t(674) = 4.65$, $p < .001$, $f = 0.18$). We also found a main effect of situated responsibility ($F(1, 674) = 3.90$, $p = .05$, $f = 0.08$) such that individuals who held positions of responsibility ($M = 3.20$, $SD = 1.78$) were more motivated to hide evidence than those in the control condition ($M = 2.93$, $SD = 1.72$). In addition, reputation-protective motivation was associated with wanting to hide evidence. Participants who were most motivated to protect group reputation were most inclined to hide incriminating evidence ($b = 0.32$, $t(674) = 8.80$, $p < .001$, $f = 0.34$). Note that none of these effects were moderated by the publicness of the transgression ($ps > 0.62$).

We then tested whether the effect of fusion on hiding evidence was mediated by concerns for reputation. The *a* path from fusion to reputation motivation ($b = 0.38$, $t(676) = 10.60$, $p < .001$) and the *b* path from reputation motivation to hiding evidence ($b = 0.30$, $t(673) = 7.54$, $p < .001$) were significant. As a result, the indirect effect of fusion via reputation motivation was significant (IE = 0.11, 95% CI = [0.08, 0.15]), and the direct effect of fusion was not significant (*c'* path; $p = .10$), which is evidence for mediation.

We also tested a parallel mediation model with situated responsibility as predictor. The *a* path from situated responsibility to reputation motivation ($b = 0.42$, $t(676) = 5.56$, $p < .001$) and the *b* path from reputation motivation to hiding evidence ($b = 0.32$, $t(673) = 8.55$, $p < .001$) were significant. The indirect effect of situated responsibility via reputation motivation was significant (IE = 0.13, 95% CI = [0.08, 0.19]), and the direct effect of situated responsibility was not significant (*c'* path; $p = .80$), providing evidence for mediation.

7.3. Discussion

Findings from Study 4 provide direct evidence for the reputational hypothesis: People who were most motivated to protect group reputation were especially likely to endorse public punishment after public, more than private, transgressions. The study also replicated Study 3's finding that people who were situationally induced to feel responsible for the group opted for public punishment more after public, relative to private, transgressions. The study also tested two alternate explanations regarding perceived immorality and group context. Consistent with work on moral psychology (Skitka, 2010; Van Bavel, Packer, Haas, & Cunningham, 2012), participants endorsed punishment more if they considered the violation to be immoral (see Table 4.2 in the SOM). However, perceived immorality was unrelated to the transgression's publicness, indicating that the effects of publicness reported in this paper were not driven by perceptions of immorality. The data also indicate that the publicness manipulation did not alter the activated group context. Finally, we were surprised that Study 4 did not replicate the interaction effect of fusion and publicness that we detected in four previous samples (Study 1, 2a, 2b, and 3). We speculate that the newly added measure of perceived immorality may have encouraged fused individuals to prioritize relatively universal moral concerns over group-related concerns.

8. Pooled analysis of effects in Studies 1–4

Collectively our four studies suggest that individuals who were strongly fused with their group or situationally induced to feel responsible for their group advocated for reputation-protective responses in the wake of moral transgressions by group members. Specifically, these individuals wanted their group to publicly punish ingroup transgressors after public, more than private, transgressions. Further, strongly fused individuals were most likely to endorse extreme, even unethical, actions intending to protect ingroup transgressors.

Some of our interaction effects were small (0.06 < = f < = 0.19), but these are consistent with effect sizes typically reported (mean $f = 0.095$ and median $f = 0.045$; Aguinis, Beaty, Boik, & Pierce, 2005), and we believe that the theoretical and practical importance of our research question overrides concerns about effect sizes. Further, noting that the interaction effect of fusion and publicness was not significant in Study 4, we conducted additional tests by pooling all our samples to make sure that our effects, even if small, were robust.

The fusion effects were tested by pooling all five samples. We tested the effects of situated responsibility by pooling just Study 3 and 4 because this variable was experimentally manipulated only in these two studies.

8.1. Methods

We pooled data from Study 1[3] ($N = 228$), Study 2a ($N = 520$), Study 2b ($N = 446$), Study 3 ($N = 936$), and Study 4 ($N = 678$) for a combined total of 2808 participants ($N_{male} = 1096$; $N_{female} = 1566$; $N_{other} = 21$; $N_{unknown} = 125$; $M_{age} = 36.20$; $SD_{age} = 12.15$, 66.40% Democrat). A sensitivity analysis revealed that our sample had 80% power to detect small interaction effects of size $f = 0.05$. To account for minor differences in materials used across studies, we standardized all dependent variables within each study before pooling. Note that including participants who were excluded in our pooled analysis did not alter our conclusions (see SOM).

8.2. Results

8.2.1. Urging the group to publicly punish the transgressive group member

8.2.1.1. Interaction of situated responsibility and publicness of transgression. A model predicting public punishment controlling for study revealed a significant interaction of situated responsibility and publicness of transgression ($F(1, 1609) = 15.69$, $p < .001$, $f = 0.10$). As shown in Fig. 7a, those who held positions of responsibility were especially likely to urge their party to publicly punish a transgressor whose transgression was public rather than private, $F(1, 1026) = 36.2$, $p < .001$. Participants in the control conditions showed no such difference ($p = .29$).

8.2.1.2. Interaction of fusion and publicness of transgression. A similar model with fusion as predictor revealed a significant interaction of fusion and publicness of transgression (β = 0.09, $t(2798) = 3.09$, $p = .002$, $f = 0.06$). As shown in Fig. 7b, strongly fused (+1 SD; β = 0.22, $t(2802) = 4.02$, $p < .001$) were especially likely to urge their party to publicly punish a transgressor whose transgression was public rather than private. Weakly fused individuals (−1 SD) showed no such difference ($p = .73$).

8.2.2. Hiding evidence incriminating the transgressive group member

We pooled standardized scores from Study 2a (i.e., destroying evidence), and studies 2b, 4, and 4 (i.e., tampering with evidence) and tested a model controlling for study. To deal with the mild positive skew of this index, we log-transformed our index before conducting analyses. Confirming our findings from preceding studies, strongly fused individuals were most likely to endorse hiding evidence incriminating ingroup transgressors (β = 0.15, $t(2571) = 7.74$, $p < .001$, $f = 0.16$). Those who were induced to feel responsible for the group were marginally more likely to hide evidence than those in the control condition ($F(1, 2571) = 3.87$, $p = .05$, $f = 0.05$). The effects of fusion and situated responsibility were not moderated by publicness ($ps > 0.20$).

[3] We included only the two conditions corresponding to private and public transgressions from Study 1, and as a result, we could use only a subset of the study's sample.

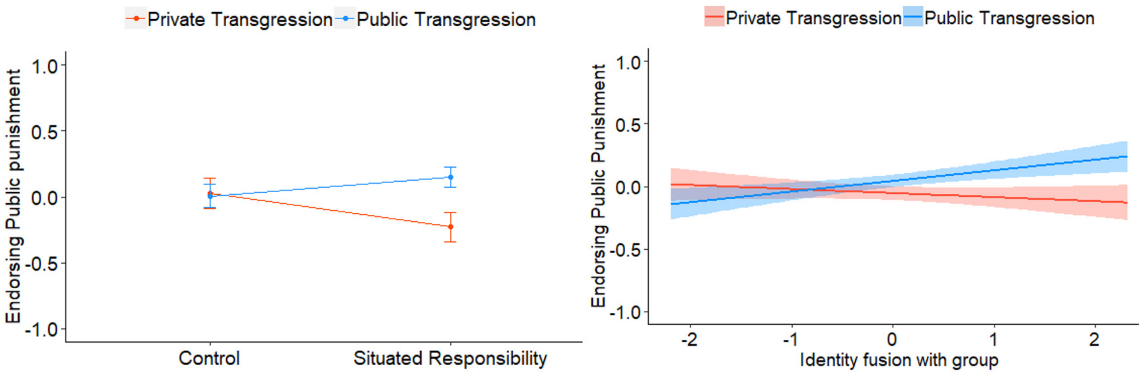

Fig. 7. Endorsing public punishment as a function of the interaction of situated responsibility (Fig. 7a) or fusion (Fig. 7b) with publicness of transgression in Study 3. Bands and error bars indicate 95% confidence intervals.

9. General discussion

In five samples with over 2800 participants, group members responded to ingroup transgressions in ways that would preserve their group's reputation. These pro-group reactions were not universal, however. Rather, reputation-protective responses were endorsed most by those who felt responsible for their group, either because they were closely aligned ("fused") with their group (Studies 1–3) or because they held positions of responsibility in their group (Studies 3 and 4). When individuals who felt responsible for the group encountered a transgression by a fellow group member, they felt motivated to protect their group's reputation (Study 4), which prompted them to opt for reputation-protective responses (Study 4). For example, after public ingroup transgressions, they symbolically distanced their group from the transgression by urging their group to publicly punish the transgressor. The same individuals, on encountering private ingroup transgressions, opted for private, as opposed to public, punishment, apparently to prevent potential reputational loss (Study 1). Further, strongly fused individuals (Studies 2–4) and those who held positions of responsibility (Study 4) were willing to contemplate unethical means of protecting the group such as destroying or tampering with evidence incriminating group members. In short, whether group members denounced or protected transgressive group members depended on whether they felt responsible for the group and on which course of action seemed most apt to safeguard the group's reputation.

Strongly fused participants presumably worked to protect the group's reputation because feelings of responsibility for the group are chronically salient for such individuals (Swann Jr et al., 2009). After public transgressions, strongly fused individuals perhaps wanted to signal their group's morality to others (Hofmann et al., 2018). On encountering private transgressions, strongly fused participants were reluctant to endorse public punishment, presumably to prevent reputational harm. Although these fusion effects failed to emerge in Study 4, the fact that they surfaced in four samples (Studies 1, 2a, 2b, and 3) and in a pooled sample of 2800 participants suggest that they are robust. Our findings suggest that similar effects would emerge with other measures of group alignment such as the indices of identification championed by advocates of social identity theory (e.g., Ellemers, Kortekaas, & Ouwerkerk, 1999; Reicher, Spears, & Postmes, 1995; Tajfel & Turner, 1979).

The current research also shows that situationally induced perceptions of responsibility for the group can influence the responses of group members to moral violations within the group. Specifically, participants who were assigned positions of responsibility resembled strongly fused group members in their zeal for wanting their group to publicly punish transgressors after public transgressions. This finding is consistent with previous reports that perceived accountability to one's group prompts pro-group behaviors (Marques, Abrams, et al., 1998). Moreover, it is possible that the effect of our modest experimental manipulation of situated responsibility is an underestimate of true effects of feelings of responsibility that accrue when people hold positions in their group for long periods of time.

Study 4 provides direct evidence for the proposed motivational mechanism regarding group reputation. Specifically, the results suggest that the reported effects of publicness are not driven by differences in perceived immorality of the transgression or the group context activated by the manipulation. Rather, it is a motivation to protect group reputation that is associated with endorsing public punishment more after public transgressions than private ones. Nevertheless, it is possible that some participants endorsed such reputation-protective actions because they believed those responses to be normative rather than because of intrinsic pro-group motivations. Knowing that other group members might reproach non-normative actions (Hornsey et al., 2005) could have inhibited their willingness to endorse actions that would threaten group reputation.

Our findings clearly have boundary conditions. Groups may sometimes prioritize concerns that are more critical or immediate than reputational concerns. For example, when a public transgression is committed by a group member who is indispensable to the survival of the group, denouncing or reporting them may be untenable because doing so would pose an existential threat to the group. In such instances, members might well cast other considerations aside and act so as to maximize the chance of their group's survival. This may explain evidence that group members apply double standards to essential figures such as leaders (Travaglino, Abrams, Randsley de Moura, & Yetkili, 2016), by displaying greater tolerance for their transgressions (Abrams, Randsley de Moura, & Travaglino, 2013). Witness, for example, the continued endorsement of eminent politicians from both prominent parties despite publicly available evidence of their moral transgressions (Edsall, 2017; Wolf, 2018).

To be sure, there is much to learn about the psychological processes that motivate group members' responses to moral transgressions within their group. Nevertheless, the current research helps illuminate the tension between tribal instincts and moral prerogatives highlighted by recent phenomena such as the #MeToo movement. The scenarios examined in the current research (i.e., transgressions within political parties) closely mirror real world events in which party members need to decide how to respond to unethical behaviors of fellow party members. Although we focused on one kind of group (i.e., political parties) and one type of moral violation (i.e., tax fraud), we suspect that our findings will generalize to other groups (e.g., universities) and transgressions (e.g., sexual assault). Future research could test the generalizability of our findings and identify other mechanisms that encourage or discourage standing up against ingroup moral violations. Ultimately such research may pave the way for the development of interventions designed to motivate people to seize the moral high ground instead of indulging their tribal instincts.

Funding

This work was supported by the National Science Foundation [grant BCS1528851 to William B. Swann, Jr.]. The funder played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Open practices

All study materials can be found in the SOM. The data used in this research has been made publicly available and can be accessed at https://osf.io/gw9jh/. The design, methods, and analysis plan of Study 2 were pre-registered, and this can be viewed at https://osf.io/eaj7f.

Acknowledgments

We thank Sanaz Talaifar for her feedback on the article, Cédric Batailler for his help with the data analysis, and PEO International for supporting Ashwini Ashokkumar with the International Peace Scholarship.

Appendix A. Supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2019.103874.

References

Abrams, D., Randsley de Moura, G., & Travaglino, G. A. (2013). A double standard when group members behave badly: Transgression credit to ingroup leaders. *Journal of Personality and Social Psychology, 105*(5), 799.

Aguiar, T., Campos, M., Pinto, I. R., & Marques, J. M. (2017). Tolerance of effective ingroup deviants as a function of moral disengagement/Tolerancia de la disidencia efectiva de los miembros del endogrupo como función de la desconexión moral. *Revista de Psicología Social, 32*(3), 659–678.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*(1), 94.

Aquino, K., Reed, A., II, Thau, S., & Freeman, D. (2007). A grotesque and dark beauty: How moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Social Psychology, 43*(3), 385–392.

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31*(2), 101–119.

Biernat, M., Vescio, T. K., & Billings, L. S. (1999). Black sheep and expectancy violation: Integrating two models of social judgment. *European Journal of Social Psychology, 29*(4), 523–542.

Bleznak, B. (2018, May 30). Ryan Gosling and More Hollywood men who have spoken out against Harvey Weinstein. *Cheatsheet*. Retrieved from www.cheatsheet.com.

Buhrmester, M. D. (2013). *Understanding the cognitive and affective underpinnings of whistleblowing* (Unpublished doctoral dissertation)Austin: University of Texas.

Buschmann, R., Henrichs, C., Pfeil, G., Windmann, A., & Wulzinger, M. (2017, April 19). Cristiano Ronaldo's Secret. *Spiegel Online*. Retrieved from www.spiegel.de.

Chavez, N., & Sutton, J. (2018, October 19). Former USA gymnastics president arrested on charge of evidence tampering in Larry Nassar case. Retrieved from www.cnn.com.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*, 377–383.

Dennis, S. A., Goodson, B. M., & Pearson, C. (2018). *Mturk Workers' use of low-cost "virtual private servers" to circumvent screening methods: A research note*.

Edsall, T. (2017, September 14). Trump says jump. His supporters ask, how high? *The New York times*. Retrieved from www.nytimes.com.

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology, 29*(2–3), 371–389.

Ellemers, N., Pagliaro, S., Barreto, M., & Leach, C. W. (2008). Is it better to be moral than smart? The effects of morality and competence norms on the decision to work at group status improvement. *Journal of Personality and Social Psychology, 95*(6), 1397.

Ellemers, N., & van den Bos, K. (2012). Morality in groups: On the social-regulatory functions of right and wrong. *Social and Personality Psychology Compass, 6*(12), 878–889.

Fredman, L. A., Bastian, B., & Swann Jr, W. B. (2017). God or country? Fusion with Judaism predicts desire for retaliation following Palestinian stabbing intifada. Social Psychological and Personality Science, 8(8), 882–887.

Gómez, Á, Brooks, M.L., Buhrmester, M. D., Vázquez, A., Jetten, J. & Swann, W. B., Jr. (2011). On the nature of identity fusion: Insights into the construct and a new measure. Journal of Personality and Social Psychology, 100, 918- 933.

Hofmann, W., Brandt, M. J., Wisneski, D. C., Rockenbach, B., & Skitka, L. J. (2018). Moral punishment in everyday life. *Personality and Social Psychology Bulletin, 44*(12), 1697–1711 0146167218775075.

Hornsey, M. J., Bruijn, P. D., Creed, J., Allen, J., Ariyanto, A., & Svensson, A. (2005). Keeping it in-house: How audience affects responses to group criticism. *European Journal of Social Psychology, 35*(3), 291–312.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1), https://doi.org/10.1146/annurev-psych-122414-033702.

Maeson & Hobson (2017, February 16). USA gymnastics says it alerted FBI to doctor accused of sex abuse in 2015. *The Washington post*. Retrieved from www.washingtonpost.com.

Marques, J., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of Personality and Social Psychology, 75*(4), 976.

Marques, J. M., Paez, D., & Abrams, D. (1998). Social identity and intragroup differentiation: The "black sheep effect" as a function of subjective social control. *Current perspectives on social identity and social categorization* (pp. 124–142). New York: Sage.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgmental extremity towards ingroup members in inter-and intra-group situations. *European Journal of Social Psychology, 18*(3), 287–292.

McGee, R. W. (2006). *The ethics of tax evasion: A survey of international business academics*.

Pagliaro, S., Ellemers, N., & Barreto, M. (2011). Sharing moral values: Anticipated ingroup respect as a determinant of adherence to morality-based (but not competence-based) group norms. *Personality and Social Psychology Bulletin, 37*(8), 1117–1129.

Pagliaro, S., Ellemers, N., Barreto, M., & Di Cesare, C. (2016). Once dishonest, always dishonest? The impact of perceived pervasiveness of moral evaluations of the self on motivation to restore a moral reputation. *Frontiers in Psychology, 7*, 586.

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology, 6*(1), 161–198.

Robinson, J. (2018, August 15). TurkPrime tools to help combat responses from suspicious geolocations. [Web log post]. Retrieved from https://blog.turkprime.com.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass, 4*(4), 267–281.

Swann, W. B., Jr., Gómez, A., Seyle, D. C., Morales, J., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology, 96*(5), 995.

Swann, W. B., Jr., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review, 119*(3), 441.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.). *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Talaifar, S., & Swann, W. B. (2019). Deep alignment with country shrinks the moral gap between conservatives and liberals. *Political Psychology, 40*(3), 657–675.

Travaglino, G. A., Abrams, D., Randsley de Moura, G., & Yetkili, O. (2016). Fewer but better: Proportionate size of the group affects evaluation of transgressive leaders. *British Journal of Social Psychology, 55*(2), 318–336.

TurkPrime (2018, September 18). After the bot scare: Understanding What's been happening with data collection on MTurk and how to stop it [web log post]. Retrieved from https://blog.turkprime.com.

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PLoS One, 7*(11), e48693.

Van der Toorn, J., Ellemers, N., & Doosje, B. (2015). The threat of moral transgression: The impact of group membership and moral opportunity. *European Journal of Social Psychology, 45*(5), 609–622.

Watt, H. (2018, March 28). Harvey Weinstein aide tells of "morally lacking" non-disclosure deal. *The Guardian*. Retrieved from www.theguardian.com.

Wolf, B. (2018, June 4). Democrats still have a Bill Clinton problem. *CNN*. Retrieved from www.cnn.com.

Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology, 115*(6), 929.

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206.