Meta-Psychology, 2023, vol 7, MP.2021.2764
https://doi.org/10.15626/MP.2021.2764
Article type: File-Drawer Report

Open data: Yes
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Rickard Carlsson
Reviewed by: Deliah Sarah Bolesta, Rima-Maria Rahal
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/VXQJ5

Group Membership and Deviance Punishment: Are Deviant Ingroup Members Actually Judged more Negatively than Outgroup Ones?

Eric Bonetto[1], Timothy S. Carsel[2], Jaïs Adam-Troian[3], Florent Varet[4], Lindsay M. Keeran[1], Grégory Lo Monaco[2], and Anthony Piermattéo[4]

[1]Aix-Marseille University Institute
[2]University of Illinois at Chicago, Chicago Illinois
[3]Department of Social Sciences, Canadian University Dubai
[4]Université Catholique de Lille, Equipe OCES

Deviance Punishment is an important issue for social-psychological research. Group members tend to punish deviance through rejection, ostracism and – more commonly - negative judgments. Subjective Group Dynamics proposes to account for social judgement patterns of deviant and conformist individuals. Relying on a group identity management perspective, one of the model's core predictions is that the judgment of a deviant target depends on group membership. More specifically, the model predicts that deviant ingroup members should be judged more negatively than outgroup ones. Although this effect has been repeatedly observed over the past decades, there is a current lack of sufficiently powered studies in the literature. For the first time, we conducted tests of Subjective Group Dynamics in France and the US to investigate whether ingroup deviants were judged more harshly than outgroup ones. Across six experiments and an internal mini meta-analysis, we observed no substantial difference in judgment between ingroup and outgroup deviant targets, d = -0.01, 95% CI[-0.07, 0.06]. The findings' implications for deviance management research are discussed.

Keywords: Deviance, Punishment, Subjective Group Dynamics, Replication

An important focus of social-psychological research about deviance pertains to the way social groups react toward members who deviate from group norms (Abrams, 2010). Group members tend to punish deviance through rejection, ostracism and more commonly through negative judgments (Bendor & Swistak, 2001; Douglas, 2010; Hogg & Reid, 2006; Lapinski & Rimal, 2005; Rimal & Real, 2003). The Subjective Group Dynamics Model (Marques, Abrams, Paez, & Hogg, 2001; Marques, Paez, & Abrams, 1998) tackles the issue of deviance punishment and predicts, amongst other things, that deviant ingroup and deviant outgroup members are not punished to the same extent.

Subjective Group Dynamics states that ingroup members will be evaluated more extremely than outgroup members because the attitudes and behaviors of ingroup members are more relevant to the ingroup's identity (Marques et al., 1998). From this perspective, pro-normative ingroup members should be evaluated more positively than pro-normative outgroup members. Conversely, deviant ingroup members should be judged more negatively than deviant outgroup members. Hence, when an ingroup member displays attitudes or behaviors that threaten the positive image of the ingroup, then other ingroup members should react negatively toward the deviant member (see the Black Sheep Effect; Marques, 2010; Marques, Yzerbyt, & Leyens, 1988; Marques & Paez, 1994). Consequently, the deviant ingroup member should be evaluated negatively or – in some cases – ostracized in an effort to restore the group's positive social identity through maintaining the perceived ingroup superiority compared to the outgroup (Marques, 2010; Abrams, Rutland, & Cameron, 2003). On the contrary, because attitudes and behaviors of outgroup members are less relevant to the ingroup's identity, reactions to a deviant outgroup member should be less extreme (Marques, 2010). Therefore, this 4 effect draws upon ingroup favoritism (i.e., the tendency to attribute more symbolic or material rewards to one's own group over an outgroup (Tajfel, Billig, Bundy, & Flament, 1971; Turner, Brown, & Tajfel, 1979), whereby individuals must reconcile their knowledge of the existence of undesirable ingroup members with their motivation to uphold a favorable view of the ingroup (Marques et al., 1988; Pinto, Marques, Levine, & Abrams, 2010).

Previous studies highlighting an effect of group membership on deviance punishment tend to present participants with a description of a target individual or a target group engaging in a deviant behavior (even if the deviance of this behavior is not systematically pretested; e.g., Wang, Zheng, Meng, Lu, & Ma, 2016). Participants are then asked to judge the target on various dimensions through self-report measures (e.g., warmth, competence).

Deviance management has attracted considerable research focus over the past decades. The effect of group membership on deviance punishment has been observed in a wide range of intergroup contexts and under a variety of collective identity threat conditions (Pinto et al., 2010; Branscombe NR, Wann DL, Noel JG, Coleman, 1993; Castano, Paladino, Coull, & Yzerbyt, 2002; Coull, Yzerbyt, Castano, Paladino, & Leemans, 2001; Hutchison & Abrams, 2003; Khan & Lambert, 1998; Shin, Freda, & Yi, 1999; Stapel, Koomen, & Spears, 1999).

As such, one might conclude that this effect is a highly replicable and robust phenomenon. In fact, it is often included in introductory psychology textbooks as an example of a robust and counterintuitive finding (Abrams, Hogg, & Marques, 2005; Albarracin, Johnson, & Zanna, 2005; Fiske, Gilbert, , & Lindzey, 2010; Levine & Hogg, 5 2010; Postmes & Jetten, 2006). Despite this amount of literature, a methodological issue suggests a call for further investigation of this effect. Many of these studies rely on small samples (e.g., N = 66 for four groups in Marques et al., 1998, experiment 1; N = 37 for four groups in experiment 2; N = 46 for two groups in Castano et al., 2002, experiment 1; N = 28 for two groups in experiment 2; see also Bettencourt et al., 2015). Because small samples are unlikely to capture extreme values that are present in the population, they tend to inflate observed effect sizes (Button et al., 2013). Consequently, it seems possible that the published effects of group membership on deviance punishment are much larger than the true effect. This problem is exacerbated as researchers conduct power analyses prior to data collection and use these inflated effect sizes because they will underestimate the number of participants they will actually need (Anderson, Kelley, & Maxwell, 2017). Therefore, these practices feed into each other and make it difficult to interpret the robustness of the effect.

The presence of this methodological issue led us to conduct a series of six sufficiently powered tests (i.e., by current, post-replication crisis standards) of Subjective Group Dynamics. More specifically, we sought to test the hypothesis according to which deviant ingroup members are punished more harshly than outgroup ones, in a time of concerns regarding the replicability of social psychological research (Earp & Trafimow, 2015; Nosek et al., 2015]. The studies reported below were conducted by independent teams in France and in the US.

Method

General Method

Over the past four years, independent teams from France and the US conducted replication attempts of the effect according to which deviant ingroup members would be evaluated more negatively than deviant outgroup members, as predicted by the Subjective Group Dynamics Model. Because the research teams were working independently of each other, our studies span several intergroup contexts and social norm violations, using different dependent variables. Consequently, our studies constitute conceptual replication attempts with samples drawn from international populations.

In each study, a deviant target was described in a vignette and participants were asked to judge this target on various dimensions (e.g., warmth, competence, social distance) through self-report measures (Branscombe et al., 1993; Khan & Lambert, 1998; Rullo, Presaghi, & Livi, 2015). All multi-item measures were mean. The present studies were conducted with the aim of achieving a sample size of at least N = 50 per condition, as recommended by Simmons, Nelson and Simonsohn (2013). After we reported the individual effects for each study, we proceeded to a mini meta-analysis of aggregated results (Goh, Hall, & Rosenthal, 2016) to try to give an estimate of the size of the effect of group membership on deviance punishment. Although some measures that were collected for exploratory purposes are not reported in this article, all data, syntaxes, supplementary information about procedures, and all measures for all studies can be found here: https://osf.io/392ha/.

All studies were conducted in accordance with the 1964 Helsinki declaration (WMO, 1964) and its later amendments, the ethical principles of the French Code of Ethics for Psychologists (CNCDP, 2012), and the 2016 APA Ethical Principles of 7 Psychologists and Code of Conduct (APA, 2017). This research was approved by the Institutional Review Board [anonymized for peer review] (Research Protocol 2017- 1027). All studies are reported, and no subject was removed from the original databases. Sample sizes for each study was determined a priori and without any extension on the basis of initial looks at the results. However, because not all participants answered every question, we used pairwise deletion on the variables for which we did not have re-

sponses. Consequently, the number of participants in each analysis fluctuates a little around the total sample size.

Details for the Six Studies

Study 1 (US, 2017)

We recruited 300 participants (60.00% male; Mage = 34.65, SD = 10.26) from Amazon's Mechanical Turk (Mturk) ($0.10/minute). A sensitivity analysis showed that this sample enabled us to detect an effect size of ρ = 0.16 at 80% power. Participants were sent a computerized questionnaire upon registering for the experiment.

Upon signing the informed consent, participants were told that they would read a short newspaper article that was printed shortly after an altercation that ostensibly happened during the 2016 Summer Olympics between the United States and Australian basketball fans. We manipulated between subjects whether the Australian fans or the United States' fans initiated the altercation. After reading the fake article, participants first completed a 2-item feelings thermometer (r = .87) about the deviant fans ('To what extent do you feel favorable and warm toward the [American fans vs. Australian fans] or unfavorable and cold toward them?' from -3 'very cold' to +3 'very warm'). Then, they filled a 2-item (r = .91) measure of blame (e.g., 'To what extent do you blame the 8 [American fans/Australian fans] for the fight between the American and Australian basketball fans?' and 'To what extent do you think the [American fans/Australian fans] are responsible for the fight between the American and Australian basketball fans?', from 1 'not at all' to 5 'very much'). Finally, participants completed a punishment measure ('To what extent do you think the [American fans/Australian fans] should be punished for their behavior?' from 1 'not at all' to 5 'very much') and provided a fine they would leverage against the deviant fans between $0 and $1000.

Study 2 (US, 2018)

Study 2 was a direct replication of Study 1 with two exceptions. First, participants were recruited via the psychology subject pool at a US University instead of through Mturk. Second, because we did not have a direct manipulation check in Study 1 on the perceived deviance of the target, we measured all variables within subjects. We also asked participants to what extent they judged the behaviors of each group of fans (Australian and United States) to be peaceable versus hostile (-3 = very hostile to +3 very peaceable), appropriate versus inappropriate (-3 = very inappropriate to +3 very appropriate), and acceptable versus unacceptable (-3 = very unacceptable to +3 very acceptable). These three items were averaged together to create the manipulation check (rUS = .89, rAustralia = .86).

We recruited 199 undergraduate students (32.00% male; Mage = 19.15, SD = 1.42). A sensitivity analysis showed that this sample enabled us to detect an effect size of ρ = 0.20 at 80% power. As in Study 1, participants completed a computerized questionnaire that was sent to them via email but instead participated in exchange for course credit. All other measures were the same: Feelings thermometer about the deviant fans rUS = .67, and rAustralia = .74; blame rUS = .80, and rAustralia = .84; punishment; and a 9 fine. The primary analyses were conducted on evaluations of the deviant fans in an independent-samples t-test, but the results do not change when analyzed as a mixeddesign (see the supplementary information on the OSF for these analysis). Therefore, these data were computed as if they came from a between-subjects design to fit with the rest of the studies.

The interaction between the instigating country and the within-subjects' evaluations of the fans on the manipulation check composite was significant, F(1, 195) = 181.48, $p < .001$, $\eta 2$ p = .48. As expected, when Australians instigated the fight, participants rated the Australians' behavior as more hostile/inappropriate/unacceptable (M = -1.72, SD = 1.29) than the American fans' behavior (M = -0.07, SD = 1.34), $t(195) = 8.93$, $p < .001$, 95% CIMD[1.29, 2.02]. When the American fans instigated the fight, participants rated the American fans' behavior as more hostile/inappropriate/unacceptable (M = -1.75, SD = 1.20) than the Australian fans' behavior (M = 0.06, SD = 1.29), $t(195) = 10.11$, $p < .001$, 95% CIMD[1.46, 2.16].

Study 3 (US, 2018)

Studies 2 and 3 were originally part of the same study. However, there was no interaction between outgroup country (i.e., Russia versus Australia) and any other independent variable. Consequently, the two conditions were separated into their own samples for ease of reporting. Please see the supplementary information on the OSF for these analyses. Study 3 was a direct replication of Study 2 with one change: Instead of the altercation between U.S. and Australian fans, the altercation was described as happening between U.S. and Russian fans. We recruited 209 undergraduate students (31.43% male; Mage = 19.04, SD = 1.18). A sensitivity analysis showed that this sample enabled us to detect an effect size of ρ = 0.19 at 80% power. As 10 in Study 2, participants completed a computerized questionnaire that was sent to them via email in exchange

for course credit. All measures were identical to those used in Study 2: Feelings thermometer rUS = .67, and rRussia = .72; blame rUS = .80, and rRussia = .83; punishment; and a fine.

As in Study 2, the same three items were used as a manipulation check (rUS = .89, rRussia = .93), and again the interaction between the instigating country and the withinsubjects' evaluation of the fans was significant, F(1, 207) = 117.04, $p < .001$, $\eta 2$ $p = .36$. When Russians instigated the fight, participants rated the Russians fans' behavior as more hostile/inappropriate/unacceptable (M = -1.58, SD = 1.43) than the American fans' behavior (M = 0.06, SD = 1.42), $t(195) = 7.90$, $p < .001$, 95% CIMD[1.23, 2.04]. When the American fans instigated the fight, participants rated the American fans' behavior as more hostile/inappropriate/unacceptable (M = -1.41, SD = 1.33) than the Russians fans' behavior (M = 0.10, SD = 1.30), $t(195) = 7.36$, $p < .001$, 95% CIMD[1.11, 1.92].

Study 4 (France, 2016)

A paper-pencil questionnaire was distributed among 143 undergraduate students (21.70% male; Mage = 19.20, SD = 1.23) in exchange for course credit. A sensitivity analysis showed that this sample enabled us to detect an effect size of $\rho = 0.23$ at 80% power. Participants were asked to read the answers of a young vs. old person target to a previous research questionnaire about homosexuality. In the 2016 European Social Survey (ESS), 88.3% of French respondents indicated that they at least 'agreed' with the statement that 'Gays and lesbians should be free to live life as they wish', indicating that homophobia is at least a somewhat deviant attitude. Analyses were conducted using the ESS online analysis tool. Weights were applied according to recommendations by the Weighting European Social Survey Data guide.

Participants were randomly assigned to one of the two vignette conditions (young ingroup member, 21 years old vs. old-outgroup member, 50 years old), and the vignette consisted of the target's answers to a few questions about their opinion about homosexuality. Participants first read the three words that came to the target's mind when we talk about homosexuality (i.e., 'problem for the society', 'pests', 'deviants'). Then, participants read the target's answers on items like 'On a scale ranging from 1 to 10, what is your opinion about homosexuals' (the responses presented the target as homophobic). Then, participants answered a 10-items ($\alpha = .94$) judgment index constructed for the study in line with the literature on social judgment (e.g., 'I have a positive image of this student', 'I think I could get along with this student', from 1 'not at all' to 8 'completely')

Study 5 (France, 2017)

An online questionnaire was distributed among social network groups (Facebook, no incentive). These social media groups were selected to be as neutral as possible, so we used trade and sales advertisements groups. We recruited 120 participants from the general French population (9.20% male; Mage = 30.61, SD = 10.86). A sensitivity analysis showed that this sample enabled us to detect an effect size of $\rho = 0.25$ at 80% power.

Participants were told that they would attend an online study about their capacity to guess the personality of others. They were asked to read interview excerpts from a target (an anonymized French vs. Belgium person) describing his/her personality (gender was not specified). The deviant targets' description was: 'After having formed an impression of something, I often find it difficult to modify. In fact, I usually do not change the way I think even after a conversation, because I have always the feeling that I'm 12 right'. In the 2016 ESS, 92.1% of French respondents indicated at least 'a little like me' to the question 'It's important to be humble and modest, not draw attention' indicating hubris and dogmatism are likely perceived to be deviant attitudes. Participants were randomly assigned to one of two vignette conditions (in-group French target vs. outgroup Belgian target). Finally, participants answered measures (7-points Likert scale, from -3 'not at all' to +3 'completely') of warmth (4 items: 'sweet', 'caring', 'amusing', 'funny'; $\alpha = .88$) and competence (4 items: 'perfectionist', 'tenacious', 'thorough', 'unshakeable'; $\alpha = .68$) for the target (Bonetto, Varet, & Troïan, 2019; Bonetto, Pichot, Girandola, & Bonnardel, 2020), and a social distance scale (Bogardus, 1933).

Study 6 (France, 2017)

Study 6 used the same deviant target description as Study 5. We recruited 161 undergraduate students (9.30% male; Mage = 20.33, SD = 3.72; no incentive). A sensitivity analysis showed that this sample enabled us to detect an effect size of $\rho = 0.22$ at 80% power. The group membership manipulation was changed to reflect the student population from which we sampled (21-years-old student target ingroup vs. 50-years-old employee target out-group), and participants were randomly assigned to one of the two vignette conditions. After reading about the dogmatic and hubristic deviant target, participants answered a 4-items ($\alpha = .83$) judgment index (e.g., 'In your opinion, X gives a good image of him/herself', from 1 'completely disagree' to 9 'completely agree'; Lo Monaco, Piermattéo, Guimelli, & Ernst-Vintila, 2011).

Results

All dependent variables were z-scored, and all analyses were independent-samples t-tests. Across all studies and all dependent measures, we did not find any evidence for an effect of group membership on deviance punishment. More precisely, we found no 13 evidence for the prediction that deviant ingroup members would be evaluated more harshly than deviant outgroup members (see Table 1). Therefore, we conducted a mini-meta analysis of our results (Goh et al., 2016) to limit the risk of making Type II errors regarding the existence of this effect in our datasets.

We meta-analyzed the results of the present six studies using the Major package for Jamovi (Hamilton, 2018). Means and standard deviations from all studies were weighted by sample size for their respective experimental group (Restricted Maximum-likelihood with mean standardized differences). The final sample size was N = 1132 (Mage = 23.83, SDage = 4.78, 27.27% male). As can be seen in Figure 1, we found no evidence for the presence of the effect of group membership on deviance punishment in our data, b = -0.01, 95% CI[-0.07, 0.06], SE = 0.03, Z = -0.15, p = .88. Effect size was d = -0.01, 95% CI[- 0.07, 0.06] and model AIC = -13.61, log-Likelihood = 8.81.

As an alternative way of conducting the meta-analysis, a mixed model was computed with dependent variables as a nested factor within studies within countries according to the following equation: judg cond + (1 | dv/study/country). Results from this analysis and scripts can be seen in the 'supplementary analysis' section (https://osf.io/5mqgw/) and converge in finding no support for substantial differences between in and outgroup deviant targets, t = 0.01, p = .99.

Discussion

This series of six conceptual replications, independently conducted by two laboratories in two different countries, did not provide evidence for the effect of group membership on deviance punishment (p = .88, d = -0.01), despite being sufficiently powered. As Earp and Trafimow (2015, p.9) put it, 'If a series of replications is carried out, independently by different labs, and deliberately tailored to the parameters and conditions so described – yet they reliably fail to produce the original result –then this should be considered informative'. These results thus provide null effects that may be used by investigators to identify boundary conditions of a deviance punishment asymmetry between ingroup and outgroup. Addressing stimulus sampling issues, we 16 tested the hypothesis across a wide range of dimensions along the attitudinal space, hence high variability across dependent variables (Fiagbenu, Proch, & Kessler, 2021; Wells & Windschitl, 1999).

Despite this variety of dependent variables, the present studies highlight consistent failures to replicate the effect of group membership on deviance punishment. Despite their consistent results, the studies also contain a number of limitations. First, an argument can be made that sufficiently powered direct replication studies are the only way to establish the presence of an effect (Doyen, Klein, Simons, & Cleeremans, 2014). However, such arguments typically fail to consider the cultural context within which an original study was conducted (Zwaan, Etz, Lucas, & Donnellan, 2018). Indeed, some experiments can be difficult, or even impossible to directly replicate (Crandall & Sherman, 2016). Consequently, a conceptual replication was the only avenue to impinge on the purported psychology.

Second, although direct replications can provide precise parameter estimates, we can never be sure that those are not artifacts due to the use of a specific paradigm and materials. Replicating an effect independent of operationalization is the only way to gain an estimate of its 'true' size, to make sure it exists as such, and that the effect is generalizable (Crandall & Sherman, 2016; Campbell & Fiske, 1959).

Third, the aggregation of such different studies (methodologically speaking) is likely to provide a biased estimate of the true effect size because of the combined noise from use of diverse methods (Huf et al., 2011). Nonetheless, the nature of the present studies allowed us to limit other typical biases found in meta analyses. All replications were conducted by independent teams (Earp & Trafimow, 2015; Berk & Freedman, 2003) 17 with different sample sizes that ranged from medium to high, which decreases the likelihood of small-study effects (Greco, Zangrillo, Biondi-Zoccai, & Landoni, 2013).

Fourth, although our results cast doubt on the claims made by previous studies regarding deviance punishment, we cannot speak to the veracity of the more general claim regarding the extremization of judgments toward ingroup targets versus outgroup targets because we focused exclusively on judgments of deviant targets. In other words, these were not replications of the well-known Black Sheep Effect (Marques et al., 1988; Marques & Paez, 1994) that were attempted here. Indeed, such an effect often refers to an interaction effect in that researchers typically manipulate whether a target is an ingroup or an outgroup member and whether the target behaves counter-normatively or pronormatively. We focused exclusively on the main effect of group membership on deviance punishment in this paper.

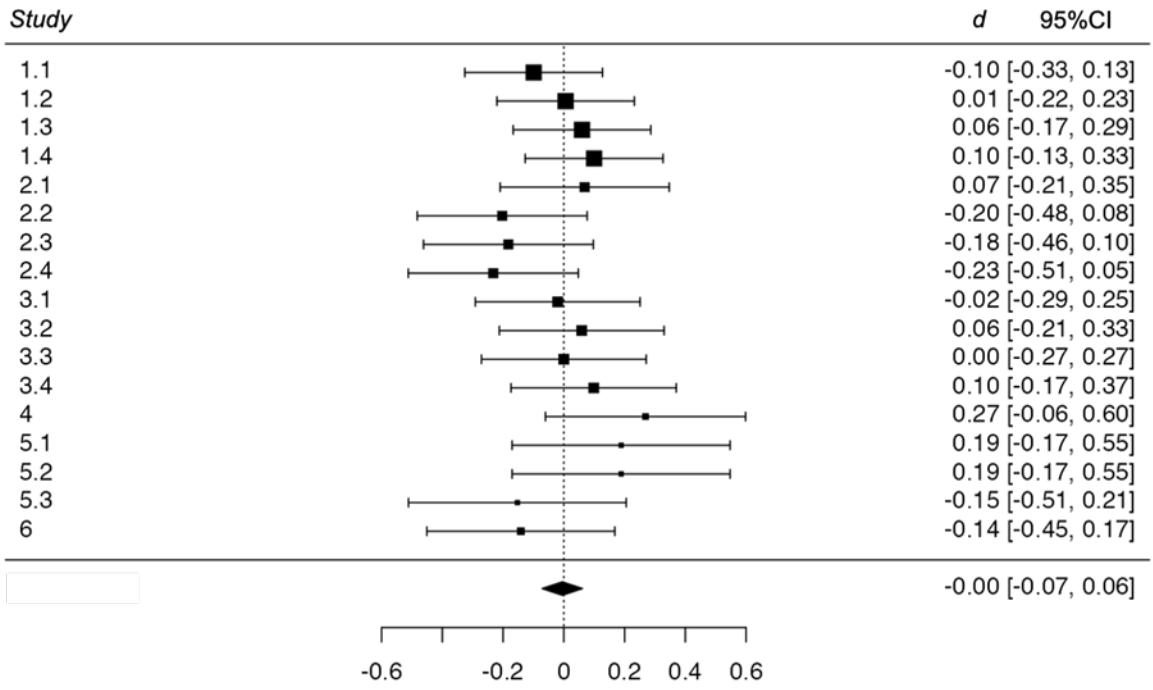
The present contribution paves the way for poten-

Table 1

Study	Measure	*N*in gr	*N*out gr	*M*ingr(*SD*)	*M*outgr(*SD*)	*t*(ddl)	*p*	*d*
1	Feeling Thermometer	151	149	0.05(1.01)	-0.47(1.00)	-0.81(298)	.42	-0.10
	Blame	151	149	-0.00(1.01)	0.00(0.99)	0.52(298)	.96	0.01
	Punishment	151	149	-0.03(1.02)	0.03(0.98)	0.52(298)	.60	0.06
	Fine	151	148	-0.05(0.95)	0.05(1.05)	0.86(297)	.39	0.09
2	Feeling Thermometer	104	95	-0.08(1.03)	-0.01(1.00)	0.53(197)	.60	0.08
	Blame	103	95	0.09(1.02)	-0.11(0.94)	-1.41(196)	.16	-0.20
	Punishment	103	95	0.06(1.03)	-0.12(0.94)	-1.30(196)	.19	-0.19
	Fine	103	95	0.08(1.06)	-0.15(0.90)	-1.62(196)	.11	-0.23
3	Feeling Thermometer	105	104	0.06(0.93)	0.04(1.04)	-0.15(207)	.88	0.02
	Blame	105	104	-0.02(0.96)	0.04(1.07)	0.42(207)	.68	0.06
	Punishment	105	104	0.03(0.91)	0.03(1.11)	0.03(207)	.98	0.00
	Fine	104	104	-0.02(0.93)	0.08(1.09)	0.70(206)	.48	0.10
4	Judgment Index	70	73	-0.14(0.83)	-0.13(1.13)	1.61(141)	.11	0.27
5	Warmth	61	59	-0.37(0.89)	-0.20(0.90)	1.04(118)	.30	0.19
	Competence	61	59	0.19(1.00)	0.37(0.89)	1.02(118)	.31	0.19
	Social Distance	61	59	0.33(1.06)	0.17(1.02)	0.85(118)	.40	-0.16
6	Judgment Index	79	82	-0.18(0.85)	-0.30(0.98)	0.86(159)	.39	-0.14

tially important theoretical advances for deviance management research in the context of intergroup processes. As Earp and Trafimow (2015) argue, null findings from conceptual replications have specific theoretical interest. Null findings of conceptual replications can establish the boundary conditions of an effect and help proponents of the theory specify under which conditions and with which kinds of materials the effect should be obtained. For instance, the effect of group membership on deviance punishment might appear only when ingroup identification is high among participants, which would be a prerequisite condition to obtain it (strength of U.S. identification was collected for Studies 2 and 3. This point was originally outside of our plan but suggested by a reviewer. Supplementary analyses indicate that the interaction between group identification and instigator on the dependent 18 variables were either not significant or in the opposite direction as predicted by Subjective Group Dynamics. See online materials for all supplementary analyses). Another methodological limitation is that our studies did not include measures of social identification with the in- and outgroup as manipulation checks. One reason for this choice is an attempt to closely replicate protocols from the literature. For instance, Marques et al. (1989) did not include any measure of social identification in their studies despite claiming a moderation by this construct. Furthermore, when social identification is indeed included, it generally taps into the ingroup only (e.g., Pinto et al., 2011), and those manipulation test do highlight that subjects display ingroup identification over and above the scale's midpoint (Pinto et al., 2011, Study 1-2) This is to be expected if not just for the fact that this identity is made salient through the survey item presence, a phenomenon at the basis of self-categorization paradigms (see Reynolds, Turner, Haslam, & Ryan, 2001). Although the absence of proper manipulation checks for social identity did not prevent researchers from routinely obtaining group membership effects, stronger tests of the theory should include those, and assess their potential moderating effect. Moreover, although previous studies highlighted a host of moderators (e.g., social identification, within-group membership status; Pinto et al., 2010; Abrams, Travaglino, Marques, Pinto, & Levine, 2018), these typically only specify when we should expect an attenuation or exac-

Figure 1



erbation of the effect. Therefore, our studies provide evidence for when a core prediction of the Subjective Group Dynamics Model might not be corroborated, and some of these well-known moderators could actually be necessary conditions for the effect studied here. Finally, as we said previously, changes in the cultural context within which the effect of group membership on deviance punishment was previously observed should also be considered. More precisely, deviance punishment may have change over time. Societal level changes may explain inconsistencies between previous studies on deviance punishment and our attempts to replicate the effect (this kind of interpretation was considered for stereotype threat; Lewis & Michalak, 2019; see also Muthukrishna, Henrich, & Slingerland, 2020).

These series of studies demonstrate that the effect of group membership on deviance punishment might be more sensitive to contextual factors than previously considered. The identification of parameter boundaries is of paramount importance for better theory specification (Earp & Trafimow, 2015). Thus, far from invalidating the basic tenets of subjective group dynamics, these results indicate that it might be a fruitful endeavor to conduct further replications of deviance management studies to clarify what these parameters are.

Author Contact

Corresponding author: Eric Bonetto (bonetto.ericbw@gmail.com).

Conflict of Interest and Funding

No conflict of interest or specific source of funding.

Author Contributions

All authors contributed equally to this research.

Open Science Practices

This article earned the Open Data and the Open Materials badge for making the data and materials openly available. The studies were not preregistered. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

Abrams D. (2010). Deviance. In J.M. Levine & M.A. Hogg (Eds.), Encyclopedia of group processes and intergroup relations (pp. 206-211). Sage.

Abrams, D., Hogg, M.A., & Marques, J.M. (2005). A social psychological framework for understanding social inclusion and exclusion. Psychology Press.

Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. Child Development, 74, 1840-1856. https://doi.org/10.1046/j.1467-8624.2003.00641.x

Abrams, D., Travaglino, G.A., Marques, J.M., Pinto, I., & Levine, J.M. (2018). Deviance credit: Tolerance of deviant ingroup leaders is mediated by their accrual of prototypicality and conferral of their right to be supported. Journal of Social Issues, 74, 36-55. https://doi.org/10.1111/josi.2018.74.issue-1/issuetoc

Albarracin, D., Johnson, B.T., & Zanna, M.P. (2005). The handbook of attitudes. Lawrence Erlbaum Associates.

Anderson, S.F., Kelley, K., & Maxwell, S.E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. Psychological Science, 28, 1547-1562. https://doi.org/10.1177/0956797617723724

Bendor J., & Swistak P. (2001). The evolution of norms. American Journal of Sociology, 106, 1493-1545. https://doi.org/10.1086/321298

Berk, R. & Freedman, D. (2003). Statistical assumptions as empirical commitments. In. T.G. Blomberg & S. Cohen (Eds.), Law, punishment and social control: Essays in honor of Sheldon Messinger (pp. 235-254). Aldine de Gruyter.

Bettencourt, B.A., Manning, M., Molix, L., Schlegel, R., Eidelman, S., & Biernat, M. (2015). Explaining extremity in evaluation of group members: Meta-analytic tests of three theories. Personality and Social Psychology Review, 20, 49-74. https://doi.org/10.1177/1088868315574461

Bogardus, E.S. (1933). A social distance scale. Sociology & Social Research, 17, 265-271.

Bonetto, E., Varet, F., & Troïan, J. (2019). To resist or not to resist? Investigating the normative features of resistance to persuasion. Journal of Theoretical Social Psychology, 3, 167-175. https://doi.org/10.1002/jts5.44

Bonetto, E., Pichot, N., Girandola, F., & Bonnardel, N. (2020). The Normative Features of Creativity: Creative Individuals are Judged to be Warmer and More Competent. The Journal of Creative Behavior. https://doi.org/10.1002/jocb.477

Branscombe, N.R., Wann, D.L., Noel, J.G., & Coleman, J. (1993). In-group or out-group extemity: Importance of the threatened social identity. Personality and Social Psychology Bulletin, 19, 381-388. https://doi.org/10.1177/0146167293194003

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

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105. https://doi.org/10.1037/h0046016

Castano, E., Paladino, M. P., Coull, A., & Yzerbyt, V. Y. (2002). Protecting the ingroup stereotype: Ingroup identification and the management of deviant ingroup members. British Journal of Social Psychology, 41, 365-385. https://doi.org/10.1348/014466602760344269

Coull, A., Yzerbyt, V.Y., Castano, E., Paladino, M.P., & Leemans, V. (2001). Protecting the ingroup: Motivated allocation of cognitive resources in the presence of threatening ingroup members. Group Processes & Intergroup Relations, 4, 327-339. https://doi.org/10.1177/1368430201004004003

Crandall, C.S. & Sherman, J.W. (2016). On the scientific superiority of conceptual replications for scientific progress. Journal of Experimental Social Psychology, 66, 93-99. https://doi.org/10.1016/j.jesp.2015.10.002

Cuddy, A. J. C., Fiske, S. T., & Glick P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. Advances in Experimental Social Psychology, 40, 61-149. https://doi.org/10.1016/S0065-2601(07)00002-0

Douglas K. (2010). Fads and Fashions. In J.M. Levine & M.A. Hogg (Eds.), Encyclopedia of group processes and intergroup relations (pp. 269-272). Sage.

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

Earp, B.D. & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. Frontiers in Psychology, 6, 621. https://doi.org/10.3389/fpsyg.2015.00621

Fiagbenu, M. E., Proch, J., & Kessler, T. (2021). Of deadly beans and risky stocks: Political ideology and attitude formation via exploration depend on the nature of the attitude stimuli. British Journal of Psychology, 112, 342-357. https://doi.org/10.1111/bjop.12430

Fiske, S.T., Gilbert, D.T., & Lindzey, G. (2010). Handbook of social psychology, Vol. 2. Wiley & Sons.

Goh, J.X., Hall, J.A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. Social and Personality Psychology Compass, 10, 535-549. https://doi.org/10.1111/spc3.12267

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. Heart, Lung and Vessels, 5, 219-225.

Hamilton, W.K. (2018). MAJOR: Meta Analysis JamOvi R. For the jamovi project. Available from: http://kylehamilton.com/#publications*selected*

Hogg M.A. & Reid, S.A. (2006). Social identity, self-categorization, and the communication of group norms. Communication theory, 16, 7-30. https://doi.org/10.1111/j.1468-2885.2006.00003.x

Huf, W., Kalcher, K., Pail, G., Friedrich, M.E., Filzmoser, P., & Kasper, S. (2011). Metaanalysis: Fact or fiction? How to interpret meta-analyses. The World Journal of Biological Psychiatry, 12, 188-200. https://doi.org/10.3109/15622975.2010.551544

Hutchison, P. & Abrams, D. (2003). Ingroup identification moderates stereotype change in reaction to ingroup deviance. European Journal of Social Psychology, 33, 497-506. https://doi.org/10.1002/ejsp.157

Khan, S. & Lambert, A.J. (1998). Ingroup favoritism versus black sheep effects in observations of informal conversations. Basic and Applied Social Psychology, 20, 263-269. https://doi.org/10.1207/s15324834basp2004$_3$

Lapinski, M.K. & Rimal, R.N. (2005). An explication of social norms. Communication theory, 15, 127-147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Levine, J.M. & Hogg, M.A. (2010). Encyclopedia of group processes and intergroup relations, Vol. 1. Sage.

Lewis Jr, N., & Michalak, N. M. (2019). Has stereotype threat dissipated over time? A cross-temporal meta-analysis. https://doi.org/10.31234/osf.io/w4ta2

Lo Monaco, G., Piermattéo, A., Guimelli, C., & Ernst-Vintila, A. (2011). Using the black sheep effect to reveal normative stakes: The example of alcohol drinking contexts. European journal of social psychology, 41, 1-5. https://doi.org/10.1002/ejsp.764

Marques, J.M. (2010). Black Sheep Effect. In J.M. Levine & M.A. Hogg (Eds.), Encyclopedia of group processes and intergroup relations, Vol. 1 (pp. 55- 57). Sage.

Marques J.M., Abrams, D., Paez, D., & Hogg, M.A. (2001). Social categorization, social identification, and rejection of deviant group members. In M.A. Hogg & R.S. Tindale (Eds.), Blackwell handbook of social psychology: Group processes, Vol. 3 (pp.400-424). Blackwell.

Marques, J.M. & Paez, D. (1994). The "black sheep effect": Social categorization, rejection of ingroup deviates, and perception of group variability. In W. Stroebe & M. Hewstone (Eds.), European review of social psychology, Vol. 5 (pp. 38-68). John Wiley.

Marques, J.M., Paez, D., & Abrams, D. (1998). Social identity and intragroup differentiation as subjective social control. In S. Worchel, J.F. Morales, D. Paez, & J.-C. Deschamps (Eds.), Social identity: International perspectives (pp. 124-142). Sage.

Marques, J.M., Yzerbyt V.Y., & Leyens, J.P. (1988). The "black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. European Journal of Social Psychology, 18, 1-16. https://doi.org/10.1002/ejsp.2420180102

Muthukrishna, M., Henrich, J., & Slingerland, E. (2020). Psychology as a historical science. Annual Review of Psychology, 72, 717-749. https://doi.org/10.1146/annurev-psych-082820-111436

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., et al. (2015). Promoting an open research culture. Science, 348, 1422-1425. https://doi.org/10.1126/science.aab2374 Pinto, I.R., Marques, J.M., Levine, J.M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? Journal of Personality and Social Psychology, 99, 107-119. https://doi.org/10.1037/a0018187

Postmes, T., & Jetten, J. (2006). Individuality and the group: Advances in social identity. Sage.

Reynolds, K. J., Turner, J. C., Haslam, S. A., & Ryan, M. K. (2001). The role of personality and group factors in explaining prejudice. Journal of Experimental Social Psychology, 37, 427-434. https://doi.org/10.1006/jesp.2000.1473

Rimal, R.N. & Real, K. (2003). Understanding the influence of perceived norms on behaviors. Communication Theory, 13, 184-203. https://doi.org/10.1111/j.1468-2885.2003.tb00288.x

Rullo, M., Presaghi, F., & Livi, S. (2015). Reactions to ingroup and outgroup deviants: An experimental group paradigm for black sheep effect. PloS One, 10, e0125605. https://doi.org/10.1371/journal.pone.0125605

Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2013). Life after p-hacking. Meeting of the Society for Personality and Social Psychology, New Orleans, LA, 17-19. https://doi.org/10.2139/ssrn.2205186

Shin, G.W., Freda, J., Yi, G. (1999). The politics of ethnic nationalism in divided Korea. Nations and Nationalism, 5, 465-484.

Stapel, D.A., Koomen, W., & Spears, R. (1999). Framed and misfortuned: Identity salience and the whiff of scandal. European Journal of Social Psychology, 29, 397-402. https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/3<397::AIDEJSP936>3.0.CO;2-6

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behaviour. European Journal of Social Psychology, 1, 149-178. https://doi.org/10.1002/ejsp.2420010202

Turner, J.C., Brown, R.J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. European Journal of Social Psychology, 9, 187-204. https://doi.org/10.1002/ejsp.2420090207

Wang, L., Zheng, J., Meng, L., Lu, Q., & Ma, Q. (2016). Ingroup favoritism or the black sheep effect: Perceived intentions modulate subjective responses to aggressive interactions. Neuroscience research, 108, 46-54. https://doi.org/10.1016/j.neures.2016.01.011

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125. https://doi.org/10.1177/01461672992512005

Zwaan, R.A., Etz, A., Lucas, R.E., & Donnellan, M.B. (2018). Making replication mainstream. Behavioral and Brain Sciences, 41, 1-61. https://doi.org/10.1017/S0140525X17001972