IIMB Management Review (2018) **30**, 77–90

available at www.sciencedirect.com

ScienceDirect

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

IIMB Management Review

Punishing an "unfair" leader: People as pragmatic politicians with in-group but fair-but-biased prosecutors with out-group

Ramadhar Singh [a,*], **Hui Koon Lim** [b], **Krithiga Sankaran** [c], **Naureen Bhullar** [c]

[a] *Amrut Mody School of Management, Ahmedabad University, Navrangapura, Ahmedabad, Gujarat, India*
[b] *Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Singapore*
[c] *Indian Institute of Management Bangalore, Bannerghatta Road, Bangalore, Karnataka, India*

Received 8 April 2015; revised 31 July 2015; accepted 10 January 2018; available online 12 February 2018

KEYWORDS
Fairness;
In-group defence;
Out-group prosecution;
Suppression;
Sequential mediation;
Punishment

Abstract Contrary to the fairness expected in the modern world, people seem to treat in-group members (us) better than out-group members (them). Do people then defend in-group members as politicians but prosecute out-group members in a fair-but-biased manner? Given information about injustices by a male or female manager, participants made outrage, attribution, attitude, and punishment responses to the manager. In-group defence held in the first three responses but fairness in punishment. However, the seeming fairness in punishment arose from bias suppression by outrage and mediation by attitude, and the order of mediation was from outrage to attitude and not vice versa.
© 2018 Production and hosting by Elsevier Ltd on behalf of Indian Institute of Management Bangalore. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

People navigate their social world by (a) using broad categories such as man *vs* woman, young *vs* old, or native *vs* immigrant, and (b) regarding the category that includes them as *in-group* (*us*) but the category that excludes them as *out-group* (*them*). One consequence of such categorisation is that the in-group, relative to the out-group, is provided with more rewards, resources, and opportunities (see, e.g., Dovidio & Gaertner (2010); Hewstone, Rubin, & Willis (2002) for reviews). Although such intergroup differentiation is widely prevalent, modern societies also censure those who appear to be blatantly unfair (Tetlock, 2002). How do people, then, resolve the conflict between the goals of favouring the in-group (Turner, Brown, & Tajfel, 1979) and also appearing as fair-minded persons (Branthwaite, Doyle, & Lightbown, 1979)? We provide a novel answer to this longstanding question, using the social-functionalist models of people as pragmatic politicians and prudent prosecutors (Tetlock, 2002).

* Corresponding author. Tel.: +917490002156.
E-mail address: singhr@ahduni.edu.in (R. Singh).
Peer-review under responsibility of Indian Institute of Management Bangalore.

https://doi.org/10.1016/j.iimb.2018.01.002
0970-3896 © 2018 Production and hosting by Elsevier Ltd on behalf of Indian Institute of Management Bangalore. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

People belong to interdependent groups. For their effective functioning in the group, they have developed accountability procedures: *Who should report to whom under what circumstances*. The cultural norms, religious scriptures, and the constitution of a nation are examples of such accountability procedures. These procedures are believed to be in the best interests of the members at large. Thus, people belonging to the collective (1) respect those procedures, (2) meet the demands of those procedures, and (3) place those demands on others. Adaptive challenges from the respective first, second, and third roles with accountability procedures turn people into *principled theologians*, *pragmatic politicians*, and *prudent prosecutors* (see, e.g., Skitka & Wisneski (2012) for a discussion).

Justice is a sacred value in modern organisations (Clay-Warner, Culatta, & James, 2013). Short-cuts with such a value in Singapore, a country known for fairness (Crisp & Hewstone, 2001; Hewstone & Ward, 1985), should, therefore, activate the theologian mindset among organisational members for whom the top priority is re-affirming that value. Nevertheless, righteous defence of sacred values gets complicated when the encroacher belongs to one's in-group (Marques, Abrams, & Serodio, 2001; Valdesolo & DeSteno, 2007) and also holds the leadership role (Abrams, de Moura, & Travaglino, 2013; Karelaia & Keck, 2013). To proclaim high standards of (i) conduct within the in-group that endows a positive social identity and (ii) fairness for the sake of personal identity (Singh, Choo, & Poh, 1998), people adopt a strategy that facilitates simultaneous pursuit of these contradictory goals. In Singh et al. (1998), for example, the goal of favouring the in-group was achieved by evaluating the in-group superior to the out-group in competence but the goal of fairness was achieved by considering them as socially equal in Singapore. Such compromise was also observed in subsequent studies in Europe (Mucchi-Faina, Costarelli, & Romoli, 2002; Mucchi-Faina, Pacilli, Pagliaro, & Alparone, 2009). The moderation of the intergroup discrimination by group status is also interpretable as a compromise strategy (van Prooijen & Lam, 2007). Because the adaptive challenges to pragmatic politicians (Tetlock, 2002) come essentially from the desire to make a positive self-presentation to others, any such compromise reflects on the politician mindset. Therefore, our central hypothesis is that people simultaneously try to defend their in-group and present themselves as fair-minded persons as if they were pragmatic politicians.

Simply demonstrating that the *in-group favouritism* prevails in some responses but *fairness* in other responses cannot be a crucial test of our hypothesis. We need to show both fairness and in-group favouritism in every response taken. If organisational injustice is an encroachment upon the sacred value by group interests, then any leader alleged of doing injustice should be punished more than the leader not alleged so. Statistically, therefore, only the main effect of injustice on punishment (see the left graph of Fig. 1) should be significant if the fairness goal alone were operative. This prediction would also come from an alternative model of people as prudent prosecutors who punish wrongdoers to uphold the normative order (Tetlock et al., 2007).

In the centre graph of Fig. 1, punishment is again shown to be higher when there is injustice than when there is no injustice. Importantly, punishment responses to the two leaders are similar in the condition of no injustice (i.e., fairness or pro-norm stance, Abrams, Marques, Bown, & Henson (2000); Marques et al. (2001)) but different in the condition of injustice (i.e., an in-group protection or out-group prosecution, Abrams et al. (2013)). Given such divergent strategies of fairness and discrimination in the same response, the Categorisation × Injustice effect should be significant. That is, social categorisation should *moderate* the effect of injustice or deviance on punishment (Abrams et al., 2013; Pinto, Marques, Levine, & Abrams, 2010; Travaglino, Abrams, de Moura, Marques, & Pinto, 2014).

People usually stereotype the out-group as *homogeneous* (i.e., they all are alike) but the in-group as *heterogeneous* (i.e., we are so different from each other, see Ostrom & Sedikides, (1992), for a review). Thus, the same information might be more important in diluting the stereotyping of the homogeneous out-group than the heterogeneous in-group (Linville, Fischer, & Salovey, 1989). In this cognitive view, the steeper slope for the out-group line than the in-group one, as in the centre graph of our Fig. 1, might not have any bearing on the motivated in-group favouritism that supposedly maintains one's positive social identity.

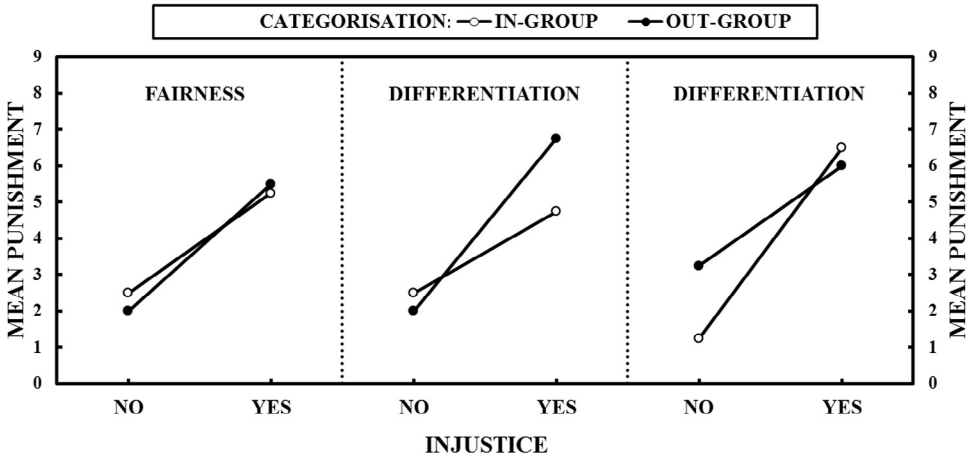


Figure 1 Predicted patterns in the Categorisation × Injustice effect by the goal of fairness and that of intergroup and intragroup differentiation.

An alternative pattern in the Categorisation × Injustice effect displayed in the right graph of Fig. 1 may, however, rekindle the motivational mechanism. Given an extreme situation of compromising the value of justice with that of group interests, injustice could evoke a uniform punitive response in both the groups. By contrast, the in-group, relative to the out-group, might be punished less or credited more for the justice done. Some evidence exists for this pattern (e.g., Travaglino et al. (2014), pp. 183–4). Such steeper slope of the in-group line than that of the out-group one is inconsistent with the cognitive view stated earlier. However, the difference at the level of no injustice is interpretable as either motivated credit to in-group leaders (Abrams et al., 2013) or fair-but-biased prosecution of out-group leaders (van Prooijen & Lam, 2007). As a sharper intragroup differentiation between normative and deviant in-group members also helps sustain the positive social identity (Marques et al., 2001; Pinto et al., 2010), this pattern is more consistent with the motivational explanation than the cognitive one.

What follows from the preceding discussion is the need to investigate subjective group dynamics in a way that differs from what has heretofore been done. In particular, the independent variables (IVs) should be the interaction between categorisation and deviance information (Abrams et al., 2013; Travaglino et al., 2014), the dependent variable (DV) should be punishment (Iyer, Jetten, & Haslam, 2012; Pinto et al., 2010; van Prooijen, 2006; van Prooijen & Lam, 2007), and the mediating variables (MVs) should be at least two other measured responses in which exactly opposite patterns, such as in the centre and right graphs of Fig. 1, might hold. Because punishment is believed to be effective in dealing with a deviant in-group member (Pinto et al., 2010, p. 115), using punishment as the DV is more justified than using anger (van Prooijen, 2006) which is immediately elicited by injustice (Adams, 1965). Thus, emotions should more appropriately be regarded as MVs of ". . . relations between appraisals of the deviant act and behavioural response such as punishment" (Otten & Gordijn, 2014, p. 175) rather than the DV. Following the prosecutorial model (Tetlock et al., 2007) and the evidence for the *ultimate attribution error* (i.e., in-group protecting but out-group derogating explanations for deviant acts (Hewstone, 1990)), we also propose that external attributions and negative attitude towards the alleged person should be conceptualised as additional MVs of punishment.

In some of the recent studies that crossed social categorisation with deviance, the hypothesised interaction effect on the responses was marginally significant even in European countries (Abrams et al., 2013, p. 804). And studies that included multiple DVs found the interaction effect on anger with, but not on favourability to, the target (van Prooijen, 2006, p. 7, p. 9) as if both the in-group favouritism and fairness goals were simultaneously pursued. Contingent upon certainty of information, moreover, the same deviant act of the in-group and the out-group produced opposite effects on punishment. In particular, the in-group, relative to the out-group, was punished more when the information about offence was certain but less when it was uncertain. To us, such divergent findings hint at prudent prosecution of a confirmed in-group offender but cracking hard on a suspected out-group offender as a biased prosecutor (Tetlock et al., 2007).

One advantage with our approach is that it allows unpacking of even the nonsignificant Categorisation × Injustice effect on punishment that otherwise denotes fairness into (i) how much of the differentiation was *suppressed* and (ii) how much of it was *mediated* by the MVs having opposite effects such as those displayed in the centre and right graphs of Fig. 1. When an MV is entered along with the IV in a regression-based mediation analysis (Hayes, 2013), the IV–DV link is reliably *increased* when the effect of a *suppressor* is controlled for but reliably *reduced* when the effect of a *mediator* is controlled for. Identifying contradictory effects arising out of the politician mindset should thus help uncover whether a nonsignificant interaction effect on the DV was "real" fairness (Branthwaite et al., 1979) or an attempt at making a positive self-presentation (Singh et al., 1998). Whereas the real fairness in the DV would remain unchanged in the moderated mediation analysis, the crafted one would change rather significantly.

Another advantage with measuring the multiple MVs is that they can specify *how* the IV effect is transmitted to the DV. In a multiple-mediator model (Hayes, 2013), the MVs may operate in *parallel*, independent of each other, or in *sequence*, that is, the preceding MV may also impact the succeeding one (Singh et al., 2015). The sequential mediation model examines the indirect effect (IE) of the IV via one MV while controlling for the IEs via the alternative MVs. Therefore, it determines which MV might be *distal* from and *proximal* to the DV, an issue that has remained ignored heretofore in the literature. Although the Group identification × Moral superiority effect on the perceived damage to group, attitude towards the target (i.e., evaluation), and punishment responses were uniform and significant, for example, Iyer et al. (2012) did not treat the first two responses as either parallel or sequential MVs in the moderated-mediation analysis of punishment. Travaglino et al. (2014, p. 184) used prescriptive focus (i.e., the extent to which the target broke the rule) as the MV for the Categorisation × Disloyalty effect on the DV of attitude towards the target. To us, both of these responses might be sequential MVs of the IV effects on the DV of punishment which was not unfortunately measured.

One might argue that our interpretation of the nonsignificant Categorisation × Injustice effect on punishment as reflecting fairness, instead of the intergroup and intragroup differentiation (Abrams et al., 2000; Marques et al., 2001), can be the inadequacy of such a two-way interaction effect. Although some of the two-way interaction effects were also nonsignificant in recent studies, the three-way interaction effect was always significant and the simple effects in them supported the hypothesised intragroup differentiation (Abrams et al., 2013, p. 112; Travaglino et al., 2014, p. 182). Because intergroup biases are often clearer in implicit than explicit measures (Crisp & Hewstone, 2001; Greenwald, Poehlman, Uhlmann, & Banaji, 2009), our prediction of a stronger two-way interaction effect in the latent MVs than in the overt DV weakens such reservation.

To demonstrate that people also take a prosecutorial stance in intragroup and intergroup differentiation but distort overt responses as do politicians, we crossed procedural injustice ($P_{Injustice}$, i.e., unfairness in procedures by which decisions are made or implemented, e.g., Folger (1977)) with distributive injustice ($D_{Injustice}$, i.e., unfairness in distribution of outcomes, resources, or opportunities, e.g., Adams (1965)). Given the importance of punishment in preserving the in-group, rather than the out-group, standard of conduct (Pinto et al.,

2010), either injustice should be sufficient for dealing with the in-group wherein morality is a distinct and important value but not with the out-group wherein morality and sociability are diffused (Leach, Ellemers, & Barreto, 2007). Also, prosecutors punish the accused for the current breach of the norm along with the past unpunished one (Goldberg, Lerner, & Tetlock, 1999), and the out-group is punished more than the in-group in uncertain circumstance (van Prooijen, 2006). The effect of one injustice can thus be amplified by another injustice in case of the out-group, a biased prosecution. If so, there should be the Categorisation $\times$ $P_{Injustice}$ $\times$ $D_{Injustice}$ effect in at least one of the MVs. Importantly, such three-way interaction in the MV(s) should again suppress and/or mediate the otherwise nonsignificant three-way interaction in the DV, showing that the intragroup and intergroup differentiation arose from the respective mindsets of prosecutors and politicians rather than the seeming fairness.

In the experiment reported, therefore, we tested five specific predictions derived from our central hypothesis:

1. Outrage, external attribution, unfavourable attitude, and punitive responses to leaders should be distinct constructs.
2. Intergroup and intragroup differentiation should manifest more vividly in MVs of outrage, attribution, and attitude than in the DV of punishment.
3. The seeming fairness in punishment should arise from suppression of the intergroup and intragroup differentiation by outrage but mediation by attribution and attitude.
4. The differentiation triggered by the IV may flow sequentially from outrage to cognitive responses and eventually to punishment.
5. The three-way interaction in any MV resulting from divergent views on morality of the in-group and the out-group should also suppress and/or mediate the otherwise nonsignificant three-way interaction in punishment.

Method

Participants and design

One hundred and twelve undergraduate students (56 males, 56 females) from the Temasek Junior College in Singapore participated in response to an appeal by the college principal. We randomly assigned the participants from each gender group to one of the eight conditions ($ns = 7$) generated by a 2 (leader's categorisation: in-group (0) *vs* out-group (1) by gender) $\times$ 2 ($P_{Injustice}$: No (0) *vs* Yes (1)) $\times$ 2 ($D_{Injustice}$: No (0) *vs* Yes (1)) between-participants factorial design. The digit in the parenthesis beside the level of a factor is its corresponding code used in data analysis.

We created the foregoing eight conditions by manipulating the three factors of social categorisation, $P_{Injustice}$, and $D_{Injustice}$ across vignettes. The vignettes stated that a *male* or *female* manager of a medium-sized private software company in Singapore was charged with making an important decision on human resource. The tasks for the manager were to determine (1) pay scales for the rest of the organisation and (2) the value of all employees' contributions to the organisation. The manager came up with the benefits and wage plans that either promoted *meritocracy* (i.e., No $D_{Injustice}$ with either male or female employees) or *favoured the in-group* by gender (Yes $D_{Injustice}$). Before making the recommendation to the management, the manager electronically shared the proposed plans with everyone on Friday afternoon, inviting comments and suggestions within either *2 weeks* (i.e., No $P_{Injustice}$ given enough time) or *3 days* (i.e., Yes $P_{Injustice}$ due to too short time and that is also over weekend). Subsequent to implementation of the recommendations, the morale in the organisation suffered. In particular, all employees expressed unhappiness with the new system.

Specifying the gender of the manager in the vignette allowed the manipulation of the social categorisation of the leader such that a same-gender manager was in-group, but an opposite-gender manager, was out-group for the participants. The earlier evidence for fairness might have been because of categorisation by race (Hewstone & Ward, 1985) which made Chinese and Malays as the majority and minority groups, respectively, in Singapore (Singh, Yeoh, Lim, & Lim, 1997; Singh et al., 1998). To eliminate such confound between categorisation and numerical status, we manipulated categorisation by gender consisting of about equal number of males (49.3%) and females (50.7%) in Singapore.[1] Further, the motive behind the seeming injustice remained ambiguous to activate prosecution of the out-group but protection of the in-group (van Prooijen, 2006). Given the evidence for a harsher punishment for the leader than a member of a group (Abrams et al., 2013; Tetlock et al., 2007), we studied injustices by the leader alone. This decision was also justified by a greater coverage of wrongdoings by leaders than members of companies in contemporary media across the globe (Karelaia & Keck, 2013).

Given seven male and seven female participants per cell, the main design was a 2 (gender of the participants) $\times$ 2 (categorisation) $\times$ 2 ($D_{Injustice}$) $\times$ 2 ($P_{Injustice}$) between-participants factorial. In an earlier survey, participants had responded to two 9-point items of group pride (i.e., *I am proud to be a man/woman; I am glad to be a man/woman*), anchored by 1 (*not all*) and 9 (*extremely*). On this measure (Spearman–Brown = .82), male ($M = 8.15$, $SD = 1.22$) and female ($M = 8.13$, $SD = 1.25$) participants scored highly but equally, $t(110) = 0.08$, $p = .94$. So, both groups were high identifiers who usually protect moral breach by members of their supposedly superior groups (Iyer et al., 2012). Consistent with the literature (Clay-Warner et al., 2013), males and females responded alike to the three manipulated factors. Thus, we excluded the factor of gender of the participants from the design ($ns = 14$ per cell).

Response measures

We measured outrage by soliciting responses to six questions of how *angry, disgusted, embarrassed, mad, turned off*, and *pained* the participants felt with the manager. We also asked them to indicate how likely it was that the manager's recommendation was caused by *chance, complexity of the task, or problems within the organisation itself*. These responses assessed external attribution to the leader's recommendation. To measure attitude towards the manager, we asked how likely it was that the participant would *vote for*

[1] http://databank.worldbank.org/data/views/reports/tableview.aspx?isshared=true.

the continuity of, *defend the style of*, and *enjoy working with* the manager. Punishment for the leader was assessed by asking how much the participant would wish that *formal complaints be made against the manager*, *he or she be made to suffer physically*, and . . . *removed from the position power*. Each judgment was made along 9-point Likert-type scales, anchored by 1 (*not at all*) and 9 (*almost certainly*). Outrage and punishment were negative responses; attribution (i.e., high score = more external explanation) and attitude (i.e., high score = favourable inclination) were, in contrast, positive responses. We reverse-scored the responses to the attribution and attitude items to ensure a uniform direction on all four measures. Thus, scores ranged from 1 (*lowest*) to 9 (*highest*).

Procedure

In a study of reactions to reports of organisational demoralisation, participants read one of the aforementioned eight vignettes in English, distributed randomly among them, and made 23 judgements about the manager. Of those judgements, 15 were of interest; eight were fillers. Responses were anonymous.

Participants read the given vignette and responded to the response measure. They worked at their own pace, and completed the task within 20 minutes. We ended each session with a full debriefing.

Results

Prediction 1

To test a four-factor structural model for the 15 responses to the leader, we performed a confirmatory factor analysis (CFA) in AMOS, with correlations among the factors. The fit of the hypothesised measurement model to the data was good, $\chi^2(84) = 146.42$, $p < .001$, non-normed fit index/Tucker–Lewis Index (NNFI/TLI) = .95, incremental fit index (IFI) = .96, root mean square error of approximation (RMSEA) = .08, standardised mean root residual (SRMR) = .06. Constraining the responses to a single factor in an alternative measurement model yielded a much worse fit to the same data, $\chi^2(90) = 446.61$, $p < .001$, NNFI/TLI = .71, IFI = .75, RMSEA = .19, SRMR = .10. The smaller χ^2 for the four-factor model than that for the alternative model, $\chi^2_{\Delta}(6) = 300.19$, $p < .001$, supported Prediction 1.

Reliability and correlation coefficients

We checked reliability of the responses constituting the foregoing four factors by Cronbach's alpha (α). The αs of the outrage, less likelihood of external attribution, unfavourable attitude, and punishment responses were .90, .63, .93, and .94, respectively. Thus, we averaged the responses to the corresponding items to form the four separate measures.

We report correlations among the four responses to the leader in Table 1. The correlations range from small to medium, justifying our use of punishment as the DV and the rest as the MVs between the IVs and the DV.

Table 1 Correlations among the four responses to the leader.

	Responses		
Responses	Attribution	Attitude	Punishment
Outrage	.25**	.58**	.49**
Attribution		.37**	.36**
Attitude			.69**

Note. $df = 110$, $^{**}p < .01$.

Prediction 2

Prediction 2 implies (i) no intergroup differentiation in punishment response but (ii) differentiation in the other three responses. To evaluate the first implication, we performed a $2 \times 2 \times 2$ ANOVA on punishment. The main effects of $P_{Injustice}$ and $D_{Injustice}$ were statistically significant. Punishment was harsher at the Yes ($M = 4.78$, $SD = 2.34$) than No ($M = 3.94$, $SD = 2.12$) level of $P_{Injustice}$, $F(1, 104) = 5.06$, $p = .03$, $\eta^2_p = .05$. Likewise, punishment was stronger at the Yes ($M = 5.37$, $SD = 2.03$) than No ($M = 3.35$, $SD = 2.03$) level of $D_{Injustice}$, $F(1, 104) = 29.24$, $p < .001$, $\eta^2_p = .22$. Notably, an unfair, compared to a fair, leader was punished more severely. However, neither main effect was moderated by the leader's categorisation. The Categorisation $\times$ $P_{Injustice}$, $F(1, 104) = 0.06$, $p = .81$, $\eta^2_p = .00$, Categorisation x $D_{Injustice}$, $F(1, 104) = 2.70$, $p = .10$, $\eta^2_p = .03$, and Categorisation $\times$ $P_{Injustice}$ $\times$ $D_{Injustice}$, $F(1, 104) = 2.91$, $p = .09$, $\eta^2_p = .03$, effects were all statistically nonsignificant. There was no main effect of categorisation either, $F(1, 104) = 1.51$, $p = .22$, $\eta^2_p = .01$. So, the results supported the first implication of Prediction 2.

Prediction 2 also implies that the main effects of injustices on the MVs should be moderated by categorisation of the leader. The effect of $D_{Injustice}$ was significant for outrage, $F(1, 104) = 83.57$, $p < .001$, $\eta^2_p = .45$, attribution, $F(1, 104) = 18.43$, $p = .001$, $\eta^2_p = .15$, and attitude, $F(1, 104) = 66.11$, $p < .001$, $\eta^2_p = .39$.[2] In ANOVAs of the MVs, moreover, the Categorisation $\times$ $D_{Injustice}$ effect was significant for outrage, $F(1, 104) = 8.75$, $p = .004$, $\eta^2_p = .08$, attribution, $F(1, 104) = 4.94$, $p = .03$, $\eta^2_p = .05$, and attitude, $F(1, 104) = 7.28$, $p = .008$, $\eta^2_p = .07$. These interaction effects indicate that the MVs, relative to the DV, were more susceptible to intergroup biases as we predicted.

We present the profile of the Categorisation $\times$ $D_{Injustice}$ effect on the MVs of outrage, attribution, and attitude and the DV of punishment across the four graphs of Fig. 2. As can be seen, the pattern in the two-way effect on outrage matches with that in the right graph of Fig. 1, reflecting on a sharper intragroup differentiation in the in-group than the out-group. In contrast, the pattern in the two-way effect on attribution and attitude matches with that in the centre graph of Fig. 1, reflecting on a harsher response to the out-group than the in-group at the Yes level of injustice. As the interaction effect on punishment was statistically nonsignificant, the pattern in the right graph of Fig. 2 can at best be regarded

[2] The main effect of $P_{Injustice}$ on the three MVs was nonsignificant, $Fs(1, 104) \leq 3.08$, $ps \geq .08$, $\eta^2_p \leq .03$. Even when people might not be subjectively bothered by $P_{Injustice}$, they might show concern for it overtly in punishment.

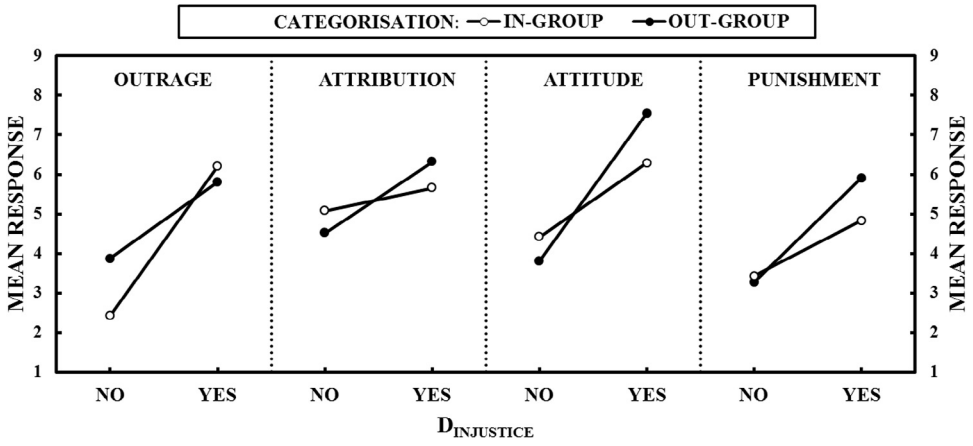


Figure 2 Mean responses to an in-group *vs* out-group leader (represented by separate lines) doing No and Yes $D_{Injustice}$ (listed on the horizontal axis).

as similar to that in the left graph of Fig. 1, illustrating fairness. Such inconsistency in the responses to the same leader is more encouraging for our social-functionalist models than the model of subjective group dynamics (Abrams et al., 2000; Marques et al., 2001).

Prediction 3

Is the foregoing inconsistency in the Categorisation × $D_{Injustice}$ effects on the DV and the MVs an outcome of the opposite causal effects of the three MVs? To answer this question, we tested simple effects of one factor at the levels of another factor of each of the four two-way interaction effects displayed in Fig. 2. The smaller the simple effect of one factor at a particular level of another factor, the greater is the importance of that particular level in determining the response (Singh, 2011). In Table 2, we first report the four means forming the two-way interaction effect on outrage, attribution, attitude, and punishment from top to bottom. For each interaction effect, we then report results from tests of significance of difference between two column means (i.e., simple effects of $D_{Injustice}$ at the two levels of categorisation) and between two row means (i.e., simple effects of categorisation at the two levels of $D_{Injustice}$). Three results from Fig. 2 and Table 2 are notable.

First, the simple effects of $D_{Injustice}$ on attribution, attitude, and punishment were stronger for an out-group than in-group leader, replicating the finding of simple *vs* complex cognitive representation of the out-group *vs* the in-group (Linville et al., 1989). On the contrary, the simple effect of $D_{Injustice}$ on outrage was much stronger for an in-group than out-group leader as if intragroup differentiation is more useful in dealing with the in-group than the out-group (Marques et al., 2001; Pinto et al., 2010).

Second, the simple effects of categorisation on attitude and punishment were present at the Yes level but virtually absent at the No level of $D_{Injustice}$. Notably, unfavourable attitude and punishment responses were higher for an out-group than in-group leader. This pattern of difference in either attitude or punishment is opposite of that in outrage. Even when there was no $D_{Injustice}$, outrage was higher for an out-group

Table 2 Means, standard deviations (SDs), and simple effects in the Categorisation × $D_{Injustice}$ effects on the four responses at the two levels of each factor.

	$D_{Injustice}$		Simple effects		
Categorisation	No	Yes	$F(1, 52)$	p	η^2_p
	Outrage				
Out-group	3.87	5.81	17.04	.001	.25
	(2.09)	(1.33)			
In-group	2.42	6.21	83.35	.001	.62
	(1.66)	(1.48)			
$F(1, 52)$	8.47	1.13			
P	.005	.29			
η^2_p	.14	.02			
	Attribution				
Out-group	4.52	6.32	18.58	.001	.26
	(1.36)	(1.76)			
In-group	5.08	5.66	2.50	.12	.05
	(1.29)	(1.41)			
$F(1, 52)$	2.47	2.50			
p	.12	.12			
η^2_p	.05	.05			
	Attitude				
Out-group	3.81	7.54	68.75	.001	.57
	(1.95)	(1.48)			
In-group	4.42	6.29	12.86	.001	.20
	(2.34)	(1.91)			
$F(1, 52)$	1.33	7.98			
p	.25	.007			
η^2_p	.03	.13			
	Punishment				
Out-group	3.27	5.91	24.75	.001	.32
	(1.90)	2.09)			
In-group	3.43	4.83	7.11	.01	.32
	(2.19)	(1.84)			
$F(1, 52)$	0.08	4.22			
p	.77	.05			
η^2_p	.00	.08			

Note. The value in parenthesis below the mean is the corresponding *SD*.

than in-group leader.[3] These results converge in suggesting that the interaction effects in outrage and attitude could be the respective suppressor and mediator of the nonsignificant Categorisation × $D_{Injustice}$ effect on punishment reported earlier.

Finally, the patterns of simple effects in attribution are not as orderly as were in those in attitude or punishment. Nonetheless, there is intergroup differentiation in this measure also. Explanations for both Yes and No $D_{Injustice}$ by an in-group leader were no different. By contrast, explanation for $D_{Injustice}$ by an out-group leader was not as less external as it was for No $D_{Injustice}$. Put simply, fairness, relative to in-group favouritism, by the out-group was more due to external factors, a kind of the ultimate attribution error (Hewstone, 1990).

Clear evidence for Prediction 3 can come from only mediational analyses that consider the MV and the DV together (Hayes, 2013). For this purpose, we first centred the three categorical IVs of leader's categorisation, $P_{Injustice}$, and $D_{Injustice}$ to their respective mean of .5 and then took their products to make seven centred terms (i.e., three main effects, three two-way interaction effects, and one three-way interaction effect). In three single-MV moderated-mediation analyses, we then specified the centred Categorisation × $D_{Injustice}$ effect as the IV, the two corresponding centred IVs as the covariates (CVs), one of the three responses of outrage, attribution, and attitude as the MV, and punishment as the DV in SPSS Process Model 4 (Hayes, 2013). The output yielded (1) the IE of Categorisation × $D_{Injustice}$ on punishment via the same interaction effect in the MV considered (IE = *ab*, where *a* = the IV effect on the MV; *b* = the MV effect when both the MV and the IV predict the DV; *c* = the total effect of the IV on the DV; and c' = the direct effect of the IV = *c*–*ab*), (2) the bias corrected 95% confidence interval (CI) around the IE from 5000 bootstrap re-samples, and (3) the original and partial effects of the two CVs. We accepted the IE as statistically different from zero only if its bias-corrected 95% CI excluded zero. We present the unstandardised regression coefficients from these analyses of the respective MVs of outrage, attribution, and attitude in the top, centre, and bottom path diagrams of Fig. 3, and the IEs via each of the three MVs in the top part of Table 3.

Three results are noteworthy. First, the path coefficients of *a* and *b* in each diagram of Fig. 3 were significant. Second, the IEs via outrage, attribution, and attitude were different from zero (see top part of Table 3). Finally, outrage turned the statistically nonsignificant interaction effect on punishment, $t = 1.61$, $p = .11$, into a significant one, $t = 2.69$, $p = .008$ (see the *c* and *c'* in the top diagram of Fig. 3). Attribution, $t = 1.16$, $p = .25$, and attitude, $t = 0.19$, $p = .85$, by contrast, further reduced the value of *c*s (see the centre and bottom diagrams of Fig. 3). Evidently, the interaction effect in outrage suppressed the corresponding interaction effect on punishment, but that in attribution or attitude mediated the very interaction effect in punishment. Thus, the nonsignificant interaction effect in punishment arose from suppression of the intergroup differentiation by outrage but mediation of it by attribution or attitude instead of fairness. So, Prediction 3 was supported.

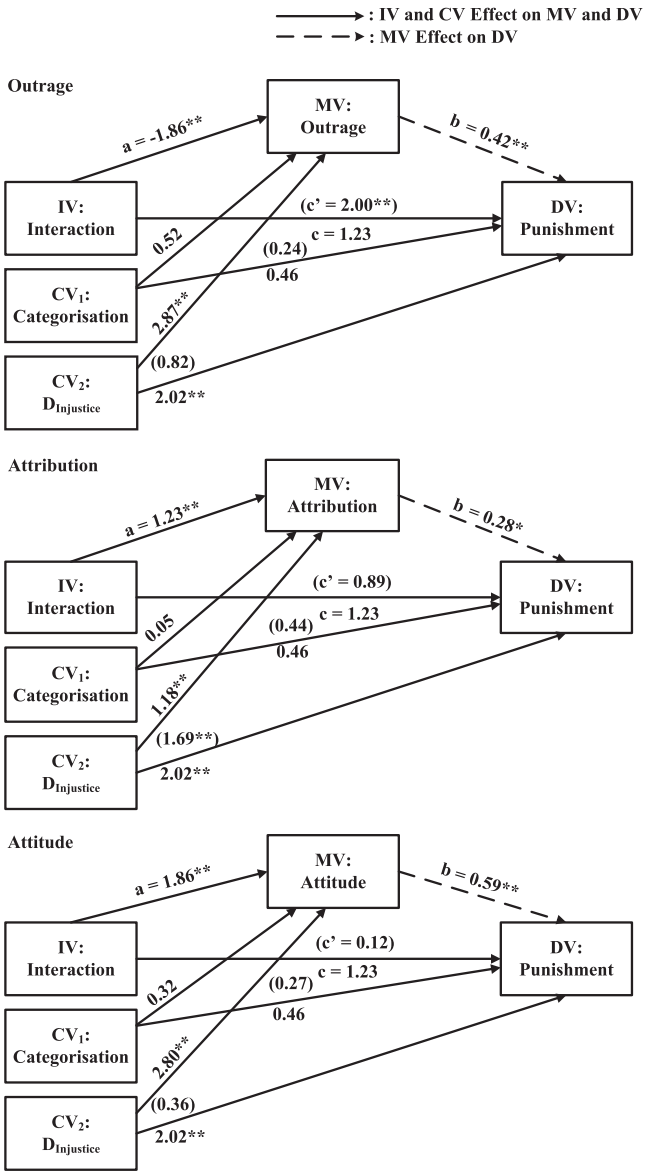


Figure 3 The unstandardised regression coefficients from three separate moderated-mediation analyses for the nonsignificant Categorisation × $D_{Injustice}$ effect on punishment, using the centred Categorisation × $D_{Injustice}$ effect in outrage (top diagram), attribution (centre diagram), or attitude (bottom diagram) as the mediator. *$p \leq .05$, **$p \leq .01$.

Prediction 4

We first tested a parallel model by entering the three MVs in a Process Model 4 analysis. Again, the IV and CVs were the same as in the single-MV analyses of Fig. 3. This analysis yielded IEs via the three MVs and the 95% CI for the differences between them. We regarded two IEs as significantly dif-

[3] A similar prosecutorial outrage with the out-group leader was suggested by a marginally significant Categorisation × $P_{Injustice}$ effect, $F(1,104) = 2.91$, $p = .09$, $\eta^2_p = .03$. At the No level of $P_{Injustice}$, outrage was significantly greater with an out-group ($M = 5.15$, $SD = 1.76$) than in-group ($M = 4.09$, $SD = 2.64$), $F(1, 52) = 7.19$, $p = .01$, $\eta^2_p = .12$, leader.

Table 3 Indirect effects (IEs) of the interaction via the corresponding interaction in the moderating variables (MVs) and their 95% confidence intervals (CIs) from tests of the mediation models.

Models		Mediators	IEs	95% CIs
Single mediator		Outrage	**−0.78**	**−0.59, −0.29**
		Attribution	**0.34**	**0.01, 0.96**
		Attitude	**1.10**	**0.24, 2.19**
Three-MV parallel		Outrage	−0.29[b]	−0.91, 0.04
		Attribution	0.19[b]	−0.04, 0.71
		Attitude	**0.96**[a]	**0.23, 1.98**
Three-MV sequential		Outrage → Attribution → Attitude		
	1	Outrage → Attitude	**−0.45**[b]	**−0.99, −0.16**
		Attitude	**1.31**[a]	**0.53, 2.31**
		Attribution → Outrage → Attitude		
	2	Outrage → Attitude	**−0.48**[b]	**−1.04, −0.19**
		Attitude	**1.31**[a]	**0.55, 2.37**
		Outrage → Attitude → Attribution		
	3	Outrage → Attitude	**−0.47**[b]	**−1.00, −0.16**
		Attitude	**1.43**[a]	**0.60, 2.54**
Two-MV sequential		Outrage → Attitude		
	1	Outrage	−0.29[a]	−0.92, 0.05
		Outrage → Attitude	**−0.48**[a]	**−1.02, −0.17**
		Attitude	**1.48**[a]	**0.65, 2.58**
		Attitude → Outrage		
	2	Attitude	**1.00**[a]	**0.23, 2.05**
		Attitude → Outrage	0.13[b]	−0.01, 0.43
		Outrage	−0.40[b]	−1.05, 0.09

Note. The IEs in bold are significantly greater than zero.
[a,b]The IEs with different superscripts differ significantly at $p = .05$.

ferent if the 95% CI of their difference excluded zero. As can be seen in the upper second part of Table 3, only the IE via attitude was different from zero.

In the next three moderated-mediation analyses by Process Model 6, we specified orders of the three MVs (Model 1: Outrage → Attribution → Attitude; Model 2: Attribution → Outrage → Attitude; and Model 3: Outrage → Attitude → Attribution). These analyses estimated dependency of the succeeding MV_2 on the preceding MV_1 (i.e., d_{21}), for example, and the IE via $MV_1 \rightarrow MV_2$ (i.e., $a_1d_{21}b_2$). Of the seven possible sources of mediation, only the two reported ones were significant in each analysis (see IEs from Models 1, 2, and 3 for the three-MVs in the third part of Table 3). Thus, we regarded outrage and attitude as potential sequential MVs but attribution as a mere correlate of punishment.[4]

In the final two moderated-mediation sequential analyses, we placed outrage as the first and the second MV to attitude. We exhibit results for such Models 1 and 2 in the top and bottom path diagrams of Fig. 4, respectively. As can be seen, the two MVs depended on each other in transmitting the IV effect to the DV, $t = 4.73$, $p < .001$. Nevertheless, the statistically nonsignificant b coefficient for outrage in both diagrams, $t = 1.51$, $p = .13$, questioned it as an MV of punishment.

The bottom part of Table 3 presents the IEs of the three sources and their 95% CIs for Models 1 and 2. As can be seen, the IE via outrage was no different from zero, but those via outrage → attitude and attitude were different from zero in Model 1. Stated simply, outrage suppressed the Categorisation × $D_{Injustice}$ effect on attitude which further mediated that effect on punishment. In Model 2, only the IE via attitude was significantly greater than zero. In both analyses, the IE via attitude was significantly greater than those via outrage, outrage → attitude, and, not or attitude → outrage which did not differ from each other. As predicted, the causal flow was from outrage to attitude and not vice versa.[5] Nevertheless, redundancy of attribution suggested a modified Prediction 4.

Prediction 5

The Categorisation × $P_{Injustice}$ × $D_{Injustice}$ effect was significant for attitude, $F(1, 104) = 14.22$, $p = .001$, $\eta^2_p = .12$. Therefore, we display the $P_{Injustice} \times D_{Injustice}$ effect on attitude towards

[4] Attribution did not mediate the main effect of $D_{Injustice}$ on punishment either, IE = 0.17, 95% CI: −0.09, 0.59.

[5] The same sequence of MVs held in the two sequential models for the main effect of $D_{Injustice}$ on punishment. The IE via outrage, 0.55, 95% CI: −0.07, 1.32, was no different from zero; those via outrage → attitude, 0.67, 95% CI: 0.31, 1.24, and attitude, 0.79, 95% CI: 0.28, 1.48, were greater than zero in Model 1. In Model 2, only the IE via attitude, 1.46, 95% CI: 0.79, 2.34, was greater than zero.

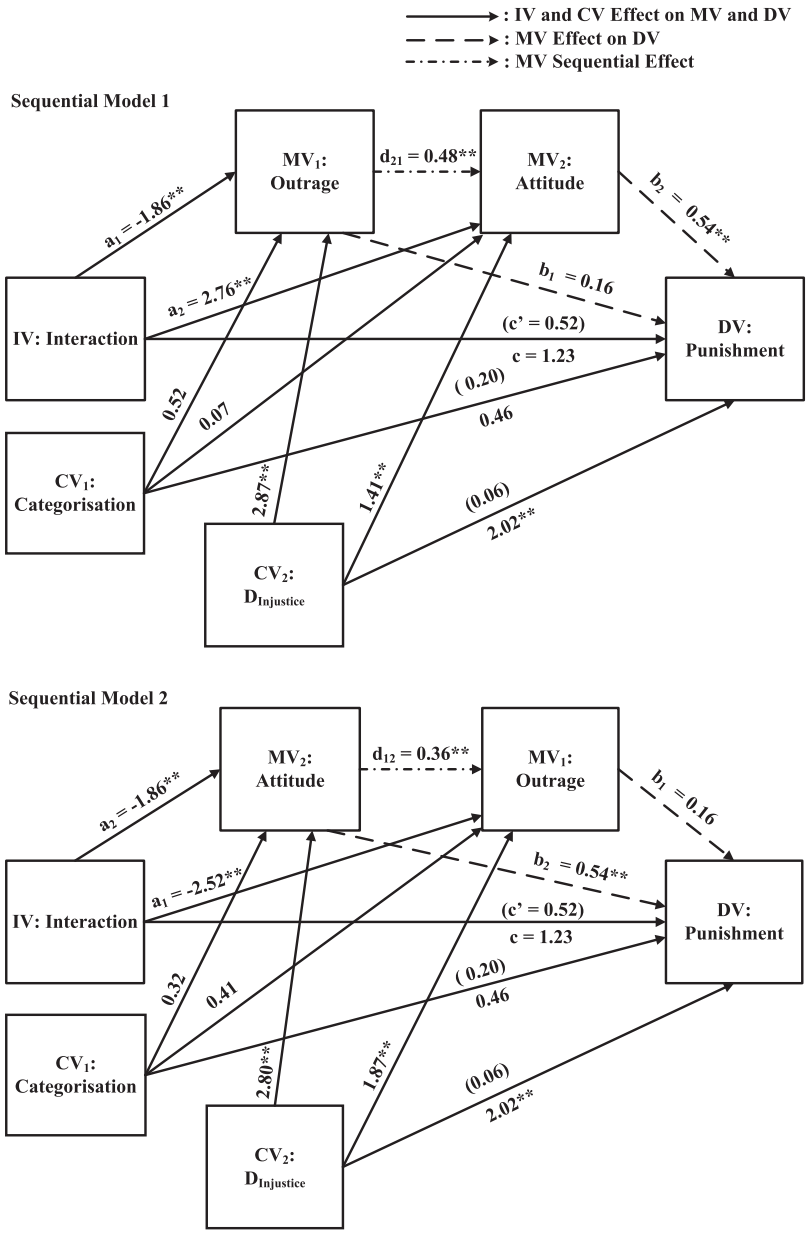


Figure 4 The unstandardised regression coefficients from two moderated-mediation analyses for the nonsignificant Categorisation × $D_{Injustice}$ effect on punishment, using the centred Categorisation × $D_{Injustice}$ effect in outrage and that in attitude as the two mediators. *$p \leq .05$, **$p \leq .01$.

the in-group and the out-group in the two left graphs and on punishment for them in the two right graphs of Fig. 5. To test Hypothesis 5, we tested simple effects of the $P_{Injustice}$ x $D_{Injustice}$ effect at their respective two levels for the in-group and out-group leaders separately. We report the four means that constituted the $P_{Injustice} \times D_{Injustice}$ effect on attitude and punishment in Table 4 and the simple effects of the interaction effect on the two responses (top and bottom parts) and for two levels of the leader's categorisation (top and bottom parts within a response).

Intragroup differentiation in processing of information about the in-group and out-group leaders is obvious in Table 4. For the in-group, the simple effect of one injustice was nonsignificant at the Yes level of another injustice but highly significant at the No level of another injustice. That is, either of the two injustices was *sufficient* for engendering an unfavourable attitude towards and a prosecutorial stance against the in-group leader. Such trend shows that injustice alone was important in fostering unfavourable attitude towards and punishment for the in-group (Singh, 2011). For an out-group leader, however, the simple effect of one injustice was stronger at the Yes than No level of another injustice. Stated differently, one kind of injustice by an out-group leader magnified the effect of another kind of injustice as if he or she was subjected to a fair-but-biased prosecution (Goldberg et al., 1999). Apparently, justice was more important than injustice in the attitude and punishment responses to the out-group. As the simple effect of $P_{Injustice}$ at the Yes level of $D_{Injustice}$

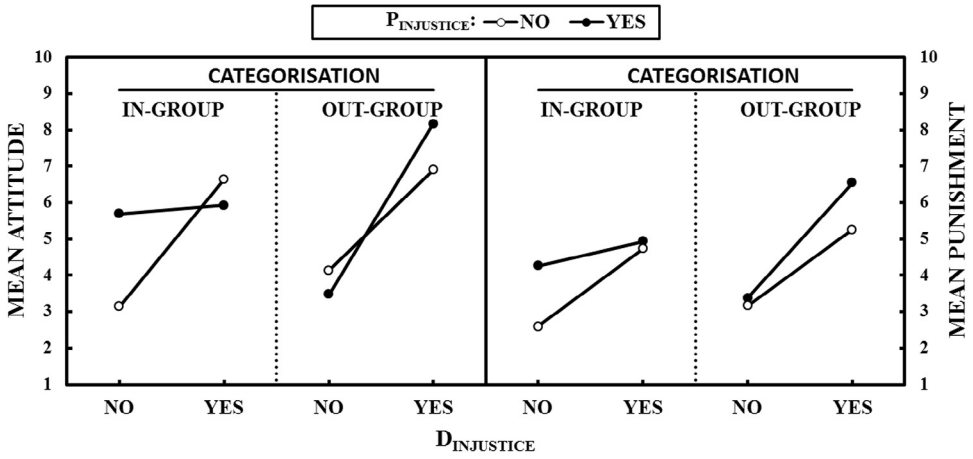


Figure 5 Mean attitude (left graphs) and mean punishment (right graphs) responses to a leader alleged to have done No or Yes $P_{Injustice}$ (represented by separate lines) and No or Yes $D_{Injustice}$ (listed on the horizontal axis). Responses to the in-group *vs* out-group leader are displayed in the respective left and right graphs of each response.

is significant for attitude but nonsignificant for punishment, distortions in overt responses were more with the out-group than the in-group (Singh et al., 1998), a result consistent with pragmatic politics.

We next did the moderated mediation analyses for the nonsignificant three-way and two-way interaction effects in punishment in the same way as for the single MV. The IV was either the centred Categorisation × $P_{Injustice}$ × $D_{Injustice}$ or Categorisation × $D_{Injustice}$ effect; the CVs were the six remaining terms; and the MV was attitude. For ease in understanding, we present regression coefficients from the analysis of the three-way interaction effect in Fig. 6. These coefficients had shifted from the CV_5 to the IV in the moderated mediation analysis of the two-way interaction effect.

Examine the coefficients for the two significant effects on attitude with the corresponding two nonsignificant ones on punishment in Fig. 6. The original total effects on punishment were substantially altered when the effect of attitude on punishment was controlled for. Further, the three-way and two-way interaction effects in attitude reliably mediated the corresponding three-way, 3.06, 95% CI: 1.41, 5.20, and two-way, 1.09, 95% CI: 0.31, 2.04, interaction effects in punishment. Interestingly, the direct effect of the three-way interaction took a negative sign, $t = -0.38$, $p = .71$, but that of the two-way interaction remained positive, $t = 0.20$, $p = .84$. On this basis, the divergent biases in attitude towards the in-group and out-group leaders reported in Fig. 4 and Table 4 can be regarded to have suppressed the three-way, but not the two-way, interaction effect in punishment. This finding further confirms the results of Fig. 3 that the suppressor of the Categorisation × $D_{Injustice}$ effect on punishment was outrage, not attitude. Thus, Prediction 5 was supported.

Discussion

Key findings

There are four key contributions. First, the outrage, attribution, attitude, and punishment responses to an erring leader are empirically distinct constructs, not interchangeable indicators of the same construct of retributive justice (van Prooijen, 2006). Evidence for such construct distinction came from the CFA of responses to the items envisaged as such and from the different patterns in the IV effects on those distinct responses.

Second, the MVs, compared to the DV, are indeed more susceptible to the intergroup and intragroup differentiations. As predicted, social categorisation of the leader seemingly moderated the effect of injustice on the MVs, but not on the DV. So, inconsistent responses to the same IV confirmed the previous finding of a compromise between the in-group favouritism and fairness goals in intergroup relations (Mucchi-Faina et al., 2002, 2009; Singh et al., 1998) and extended it to punishment of leaders (Abrams et al., 2013).

Third, the seeming "fairness" in punishment as evinced by only the two statistically main effects of $P_{Injustice}$ and $D_{Injustice}$ on punishment (Branthwaite et al., 1979) was not as such, but a result of opposing causal effects of outrage and attitude on the punishment eventually recommended. As predicted, the respective Categorisation × $D_{Injustice}$ effects on outrage and attitude had suppressed and mediated the nonsignificant Categorisation × $D_{Injustice}$ effect on punishment. Likewise, the opposite processing of the information about $P_{Injustice}$ and $D_{Injustice}$ by the in-group and out-group leaders in attitude had suppressed the Categorisation × $P_{Injustice}$ × $D_{Injustice}$ effect on punishment. Such a variety in patterns of categorisation effects across responses indicates that people do behave more like pragmatic politicians with an in-group leader but more like fair-but-biased prosecutors with an out-group leader.

Finally, outrage, attribution, or attitude per se does modify the effect size of the interaction between categorisation and injustice in punishment as if each were a reliable MV. When the three MVs were pitted together, however, attribution emerged as a correlate of punishment and outrage determined punishment only via unfavourable attitude towards the erring leader. As we showed, outrage, which was not an MV of the Categorisation × $D_{Injustice}$ effect on punishment in either three-MV parallel and sequential-mediation analyses or two-MV sequential-mediation analyses, did sequentially influence punishment through attitude. In other words, the order of mediation was from outrage to attitude and not vice versa.

Table 4 Means, SDs, and simple effects of the $P_{Injustice} \times D_{Injustice}$ interaction for the in-group and out-group leaders.

	$D_{Injustice}$		Simple effects		
$P_{Injustice}$	No	Yes	$F(1, 26)$	p	η^2_p
Categorisation $\times$ $P_{Injustice}$ $\times$ $D_{Injustice}$ effect on attitude					
	In-group				
Yes	5.69	5.93	0.09	.77	.00
	(2.18)	(1.97)			
No	3.14	6.64	26.18	.001	.50
	(1.77)	(1.85)			
F(1, 26)	11.50	0.98			
p	.002	.33			
η^2_p	.31	.04			
	Out-group				
Yes	3.48	8.17	68.39	.001	.73
	(1.85)	(1.03)			
No	4.14	6.90	15.69	.001	.38
	(2.05)	(1.61)			
F(1, 26)	0.81	6.09			
p	.38	.02			
η^2_p	.03	.19			
Categorisation $\times$ $P_{Injustice}$ $\times$ $D_{Injustice}$ effect on punishment					
	In-group				
Yes	4.26	4.93	0.56	.46	.02
	(2.63)	(2.06)			
No	2.60	4.74	14.67	.001	.36
	(1.26)	(1.67)			
F(1, 26)	4.57	0.07			
p	.04	.79			
η^2_p	.15	.00			
	Out-group				
Yes	3.38	6.55	22.53	.001	.46
	(1.96)	(1.54)			
No	3.17	5.26	6.51	.02	.20
	(1.90)	(2.41)			
$F(1, 26)$	0.08	2.82			
p	.77	.11			
η^2_p	.00	.10			

Note. The value in parenthesis below the mean is the corresponding *SD*.

Accordingly, outrage and attitude should be regarded as variables distal from and proximal to punishment, respectively.

Implications

Conceptual

Our findings re-affirm the in-group favouritism as a general norm of intergroup relations. Such a norm prevails regardless of whether the categorisation is by gender, race, or nationality (Hewstone & Ward, 1985; Singh et al., 1997, 1998). To access intergroup differentiation, therefore, one need not take implicit measures of memory (Crisp & Hewstone, 2001) or reaction time (Greenwald et al., 2009) that are under cloud lately (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Explicit paper-and-pencil measures such as those used by us can also uncover the in-group favouritism and/or the out-group prosecution if they are conceptualised and measured as the MVs of punishment.

Another conceptual implication lies in providing a new interpretation of the compromise strategy that was previously construed as indicating fairness (Mucchi-Faina et al., 2002, 2009; Singh et al., 1998). As we noted, punishment for the erring leader had only the main effects of $P_{Injustice}$ and $D_{Injustice}$. On the surface, such additive effects argue for "real" fairness: An unjust leader is punished more than the just one (Branthwaite et al., 1979). Contrary to such simplistic interpretation, the intergroup differentiation in punishment was suppressed by the opposite pattern of interaction effect in outrage. When such suppression effect was controlled for, intergroup differentiation in the overt punishment became similar to that in the latent attitude.

The use of the multiple sequential MVs also shows that the in-group favouritism serves as an *anchor* to which inconsistent but tolerable responses are *adjusted* (Epley & Gilovich, 2006) for making a positive self-presentation (Tetlock, 2002). Had justice activated an automatic generalised fairness as reflected in attitude and punishment, outrage with the out-group should not have been greater than that with the in-group. Considered together, therefore, people seem to have acted like pragmatic politicians with the in-group but fair-but-biased prosecutors with the out-group. The prosecutorial posture that they assumed with the in-group as in the latent outrage (i.e., greater intragroup differentiation with the in-group than the out-group) was eventually converted into flexibility of pragmatic politicians.

Applied

We initiated our research with the major premises that modern organisations value justice and fairness (Clay-Warner et al., 2013), and that employing people from diverse categories (van Knippenberg & Schippers, 2007), are positive steps towards further promoting this value. When mere information that the manager was a male or female can engender a complex chain of in-group protective and out-group prosecuting strategies, what might be happening to people varying in nationality or race in real and virtual organisations?

Our findings may be questioned on the grounds that college students lack clear attitudes and stable social identities and hence are easily prone to external influence (Sears, 1986) such as the information about gender of the manager. Before doing so, one should consider two other organisational phenomena, namely, (a) people who have successful careers are often similar to rather than dissimilar from their supervisors and peers (Schaubroeck & Lam, 2002), and (b) they "judge favourably those who are most similar to them" (Pfeffer, 2013, p. 275). Can there be a social category more readily visible than gender?

Given our evidence for real in-group favouritism but pseudo out-group fairness activated by the social category of gender, should the diversity programme in organisations be abandoned? We argue against gender-segregated organisations because intergroup bias is reducible by creating more opportunities for contact (see, e.g., Pettigrew & Tropp, 2006, for a review) and cross-categorisation (see, e.g., Crisp & Hewstone (2006) for a collection of chapters). An important applied implication of our findings lies in making young men and women in Singapore aware that they still discriminate between colleagues based on gender, and that they ought to

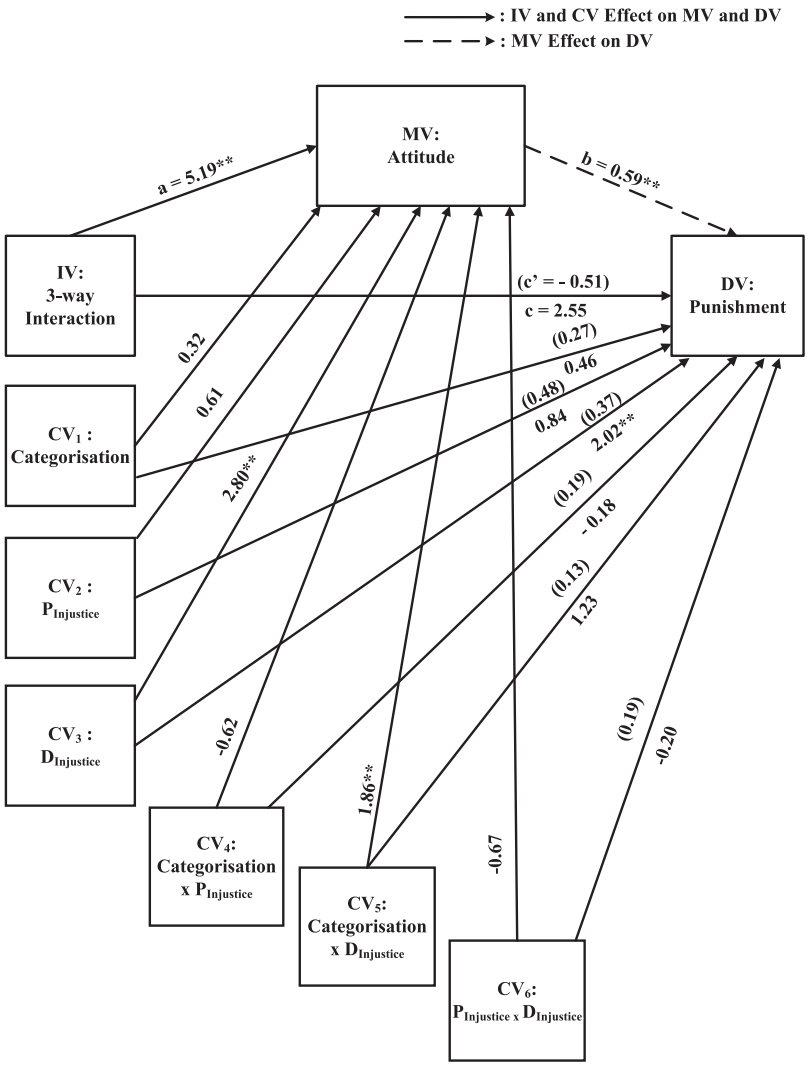


Figure 6 The unstandardised regression coefficients from the moderated mediation analysis for the nonsignificant Categorisation × $P_{Injustice}$ × $D_{Injustice}$ effect on punishment, using the centred Categorisation × $P_{Injustice}$ × $D_{Injustice}$ effect in attitude as the mediator. The coefficients remained the same when a moderated-mediation analysis was done for the Categorisation × $D_{Injustice}$ effect on punishment, using the centred Categorisation × $D_{Injustice}$ effect in attitude as the mediator. *$p \leq .05$, **$p \leq .01$.

spend additional time and effort on correcting it (Wegener & Petty, 1997). Organisations should further facilitate the bias correction by making themselves more diverse than what they currently are to reduce the saliency of gender category (Brief et al., 2005) and providing new recruits with opportunities for self-revealing similarity in peripheral attributes (e.g., attitudes, emotional experiences, interests, or values) to the opposite-gender colleagues in getting-acquainted sessions (West, Magee, Gordon, & Gullett, 2014) of the induction training and the subsequent organisational socialisation.

Limitations and future direction

Three limitations of the present research deserve mention. First, the two groups differed in attitude at the Yes level of injustice but in outrage at the No level of injustice. We interpreted the first differentiation as an in-group defence (Abrams et al., 2013) but the second one as an out-group prosecution (Singh et al., 1997). Since positive social identity can be affirmed by either the in-group favouritism or the out-group denigration (Hewstone et al., 2002), both differences might have arisen from either process. This ambiguity in the locus of the intergroup differentiation can be removed in the future research by including a control condition of no information (Singh et al., 1998) about the leader's categorisation.

Second, our manipulation of social categorisation by specifying the gender of the manager was too explicit to ignore. Thus, the obtained intergroup and intragroup differentiation might have stemmed more from the expected gender discrimination than what might really be prevalent. Had such an experimental demand contaminated the results, the differences across the four responses should have been uniform, not as divergent as found. Importantly, the sequential-mediation model in which the effect of the IV sequentially travelled from outrage to attitude and then to punishment should not have been supported so clearly either. While we

dismiss any contamination in our results, we do recommend that future experiments should use the male and female names to manipulate social categorisation by gender.

Finally, the social-functionalist models of people as theologians, politicians, and prosecutors posit causal attributions as MVs of punishment (Tetlock, 2002; Tetlock et al., 2007). Contrary to this prescription, attribution turned out to be a correlate of punishment in this research. We sought external explanations primarily because Singaporeans differed from Americans more in external than internal attributions (Tetlock, Self, & Singh, 2010). Our measure of attribution was also much less reliable than other three measures. Nevertheless, no injustice, relative to the in-group favouritism, by the out-group was attributed more to the external factors as if explanations were justifying (Alicke, 2000) rather than driving punishment (Tetlock et al., 2007). To remove this ambiguity surrounding the role of attribution in punishment, future investigators should measure dispositional instead of external attribution.

Conclusion

In responding to erring leaders, people do use double standards (Abrams et al., 2013): While people are essentially in-group favouring, they do make inconsistent responses across different measures to appear as fair-minded persons (Singh et al., 1998). When the effects in the inconsistent latent responses such as outrage and attitude are controlled for, the nonsignificant effects in punishment otherwise suggesting fairness support the motivated intragroup and intergroup differentiations in the service of maintaining a positive social identity (Otten & Gordijn, 2014). It may be correct, therefore, to regard people as pragmatic politicians with an erring in-group leader but fair-but-biased prosecutors with an erring out-group leader.

Acknowledgements

The research reported was supported by Grant R-107-000-045-112 from the National University of Singapore and the reanalyses reported in this article were facilitated by the Indian Institute of Management Bangalore Grant 7358 to the first author. An early version of this paper was presented in a symposium at the International Congress of Psychology held on July 24, 2012 at Cape Town, South Africa.

References

Abrams, D., de Moura, G. R., & Travaglino, G. A. (2013). A double standard when group members behave badly: Transgression credit to in-group leaders. *Journal of Personality and Social Psychology*, *105*, 799–815.

Abrams, D., Marques, J. M., Bown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology*, *78*, 906–912.

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). New York: Academic Press.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574.

Branthwaite, A., Doyle, S., & Lightbown, N. (1979). The balance between fairness and discrimination. *European Journal of Social Psychology*, *9*, 149–163.

Brief, A. P., Umphress, E., Burrows, J., Dietz, J., Butz, R., & Scholten, L. (2005). Communities matter: A research note on the impact of racial demographics in organizations. *Academy of Management Journal*, *48*, 830–845.

Clay-Warner, J., Culatta, E., & James, K. R. (2013). Gender and organizational justice preferences. *Sociology Compass*, *7*, 1074–1084.

Crisp, R., & Hewstone, M. (2001). Multiple categorization and implicit intergroup bias: Differential category dominance and the positive-negative asymmetry effect. *European Journal of Social Psychology*, *31*, 45–62.

Crisp, R., & Hewstone, M. (2006). *Multiple social-categorizations: Processes, models, and applications*. East Sussex. UK: Psychology Press.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, pp. 1084–1124). New York: Wiley.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*, 311–318.

Folger, R. (1977). Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequity. *Journal of Personality and Social Psychology*, *35*, 108–119.

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology*, *29*, 781–795.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.

Hewstone, M. (1990). The "ultimate attribution error"? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology*, *20*, 311–335.

Hewstone, M., & Ward, C. (1985). Ethnocentrism and causal attribution in Southeast Asia. *Journal of Personality and Social Psychology*, *48*, 614–623.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604.

Iyer, A., Jetten, J., & Haslam, S. A. (2012). Sugaring o'er the devil: Moral superiority and group identification help individuals downplay the implications of ingroup rule-breaking. *European Journal of Social Psychology*, *42*, 141–149.

Karelaia, N., & Keck, S. (2013). When deviant leaders are punished more than non-leaders: The role of deviance severity. *Journal of Experimental Social Psychology*, *49*, 783–796.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*, 234–249.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*, 165–188.

Marques, J. M., Abrams, D., & Serodio, R. G. (2001). Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. *Journal of Personality and Social Psychology*, *81*, 436–447.

Mucchi-Faina, A., Costarelli, S., & Romoli, C. (2002). The effects of intergroup context of evaluation on ambivalence toward the in-group and the out-group. *European Journal of Social Psychology*, *32*, 247–259.

Mucchi-Faina, A., Pacilli, M. G., Pagliaro, S., & Alparone, F. R. (2009). Ambivalence in intergroup evaluation: The role of fairness norm. *Social Justice Research*, *22*, 117–133.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*, 536–552.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192.

Otten, S., & Gordijn, E. H. (2014). Was it one of us? How people cope with misconduct by fellow ingroup members. *Social and Personality Psychology Compass*, *8*, 165–177.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

Pfeffer, J. (2013). You're still the same: Why theories of power hold over time and across contexts. *The Academy of Management Perspectives*, *27*, 269–280.

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.

Schaubroeck, J., & Lam, S. S. K. (2002). How similarity to peers and supervisor influences organizational advancement in different cultures. *Academy of Management Journal*, *45*, 1120–1136.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515–530.

Singh, R. (2011). Information integration as a basic cognitive process. In G. Mishra (Ed.), *Handbook of psychology in India* (pp. 73–98). New Delhi: The Oxford University Press.

Singh, R., Yeoh, B. S. E., Lim, D. I., & Lim, K. K. (1997). Cross-categorization effects in intergroup discrimination: Adding versus averaging. *British Journal of Social Psychology*, *36*, 121–138.

Singh, R., Choo, W. M., & Poh, L. L. (1998). In-group bias and fair-mindedness as strategies of self-presentation in intergroup perception. *Personality and Social Psychology Bulletin*, *24*, 147–162.

Singh, R., Wegener, D. T., Singh, S., Sankaran, K., Lin, P. K. F., Seow, M. X., et al. (2015). On the importance of trust in interpersonal attraction from attitude similarity. *Journal of Social and Personal Relationships*, *32*, 829–850.

Skitka, L. J., & Wisneski, D. C. (2012). Justice theory and research: A social functionalist perspective. In J. Suls & H. Tennen (Eds.), *The handbook of psychology* (2nd ed., pp. 407–427). Hoboken, NJ: Wiley.

Tetlock, P. E. (2002). Social-functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*, 451–471.

Tetlock, P. E., Visser, P., Singh, R., Polifroni, M., Scott, A., Elson, B., et al. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, *43*, 195–209.

Tetlock, P. E., Self, W. T., & Singh, R. (2010). The punitiveness paradox: When is external pressure exculpatory- and when a signal just to spread blame? *Journal of Experimental Social Psychology*, *46*, 388–395.

Travaglino, G. A., Abrams, D., de Moura, G. R., Marques, J. M., & Pinto, I. R. (2014). How groups react to disloyalty in the context of intergroup competition: Evaluations of group deserters and defectors. *Journal of Experimental Social Psychology*, *54*, 178–187.

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in in-group favouritism. *European Journal of Social Psychology*, *9*, 187–204.

van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515–541.

van Prooijen, J. W. (2006). Retributive reactions to suspected offenders: The importance of social categorizations and guilt probability. *Personality and Social Psychology Bulletin*, *32*, 715–726.

van Prooijen, J. W., & Lam, J. (2007). Retributive justice and social categorizations: The perceived fairness of punishment depends on intergroup status. *European Journal of Social Psychology*, *37*, 1244–1255.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*, 689–690.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). New York: Academic Press.

West, T. V., Magee, J. C., Gordon, S. H., & Gullett, L. (2014). A little similarity goes a long way: The effects of peripheral but self-revealing similarities on improving and sustaining interracial relationships. *Journal of Personality and Social Psychology*, *107*, 81–100.