



The password is praise: Content of feedback affects categorization of feedback sources

Anna Rabinovich*, Thomas A. Morton, Emily Landon, Caitlin Neill, Sapphire Mason-Brown and Lucie Burdett

University of Exeter, UK

In three experimental studies, we investigated the effect of the content of group-directed feedback on categorization of the feedback source as an ingroup or an outgroup member. In all studies, feedback valence (criticism vs. praise) and the attributional content of feedback (attributing outcomes to internal properties of the group vs. external circumstances) were experimentally manipulated. The results demonstrated that anonymous (Study 1) and ambiguous (Studies 2 and 3) sources of feedback are more likely to be seen as (typical) ingroup members when they provide praise rather than criticism. In addition, in all studies there was a significant interaction between valence and the attributional content of feedback, such that sources of praise were more likely to be seen as ingroup members when they attributed the group's success to internal (rather than external) causes, while the opposite was observed for critics. These effects were mediated by perceived group image threat. Implications for research on group-based feedback and social categorization are discussed.

Imagine reading through online comments on a newspaper article and coming across one that heavily criticizes the political party that you support. You do not know anything about the person who posted this comment, but you will probably not hesitate to draw some conclusions about their political affiliation. In fact, most of us would be likely to make such a conclusion straight away, and to consider the content of the comment through the prism of this inferred group membership. Similar processes are at play when authors receive anonymous reviews of their research. When receiving generally positive feedback about one's work, one often jumps to conclusions about the reviewer's likely affiliation to a particular shared school of thought. As most authors would also know, such attributions lay the ground for easy discounting of critical feedback.

In fact, in many instances of receiving feedback (either critical or flattering), we do not have much information about the people who provide such feedback but nonetheless make quick inferences about them. Given that information about the group membership of a feedback provider, when available, has been shown to have substantial effects on responses to feedback (Hornsey, 2005), it would seem important to also understand how such information is inferred when not provided explicitly. This research addresses this question by exploring the effect of feedback content on inferences made about either anonymous or ambiguously characterized feedback sources. In particular, we focus on

*Correspondence should be addressed to Anna Rabinovich, School of Psychology, University of Exeter, Perry Road, Exeter EX4 4QG, UK (e-mail: a.rabinovich@ex.ac.uk).

how categorization of a feedback source as ingroup or outgroup is affected by the content of their feedback. We start by reviewing previous research on responses to group-based feedback, followed by a brief discussion of previous studies on the categorization of ambiguous social stimuli. Finally, we present three empirical studies exploring categorization of ambiguous sources as a function of variations in the valence and attributional content of their feedback.

Responses to group-based feedback

The group membership of feedback sources has been shown to play an important role in determining responses to them. In particular, criticism coming from outgroup members typically evokes negativity and is more likely to be rejected than identical criticism coming from inside the group (the intergroup sensitivity effect; Hornsey & Imani, 2004). This effect is explained by the inferred motives of ingroup and outgroup critics: ingroup critics are attributed constructive motives (i.e., they are perceived to be acting in the best interests of the group) whereas outgroup critics are perceived to be driven by intergroup competition and hostility (Hornsey, Trembath, & Gunthorpe, 2004). In this research paradigm, affiliation of the critic is treated as a key factor in determining responses to feedback, and consequently such information is provided explicitly in the experimental materials.

More recent research in the domain of group-directed criticism has demonstrated that the intergroup sensitivity effect can be moderated by attributional content of criticism – that is, the reasons for failure invoked by the critic. Somewhat paradoxically, outgroup critics can sometimes be effective at eliciting remedial action, especially when they make internal (rather than external) attributions for ingroup failure, – for example, when they explain the group’s poor performance through reference to its enduring character rather than its external circumstances (Rabinovich & Morton, 2010). The surprising effectiveness of this form of outgroup criticism was demonstrated to stem from the particular threat it poses to the group’s external image – a threat that group members become motivated to defend against by refuting the criticism through their own good behaviour. Interestingly, similar results have been observed in the domain of positive feedback (i.e., praise). Specifically, recipients of praise are more likely to behave in line with this feedback when the positive image of their group is not unequivocally affirmed, – for example, when outgroup members attribute group’s success to external, rather than internal factors (Rabinovich, Morton, Crook, & Travers, 2012). Again, attributing positive performance externally was shown to threaten the image of the group and result in behaviour that re-affirms the group’s positive qualities. Conversely, internally attributed praise directly affirms the group’s image without the need for further action by group members.

In sum, research on both criticism and praise demonstrates that responses to feedback are determined not only by who is providing this feedback but also by what they are saying. Certain types of feedback (such as internally attributed criticism and externally attributed praise) may be threatening to one’s group image. Of course, it should be noted that attributions labelled as ‘external’ here, and in previous research on this topic (e.g., Rabinovich & Morton, 2010; Rabinovich, Morton, Crook, & Travers, 2012) are not completely external to the group. Instead, they refer to ingroup’s structural conditions and institutional practices (e.g., available facilities) as opposed to group’s character. Nonetheless, these patterns do show that the threat arising from specific combinations of

feedback and attribution is one key process driving individual responses to group-based feedback, especially when this feedback comes from the outside.

Although previous research shows that group membership shapes responses to criticism, this work has focussed on situations in which information about the group membership of a critic is not only available but also made quite explicit. Therefore, it is unsurprising that this parameter dominates responses to criticism and that other factors (e.g., criticism content) are often seen as secondary to, or framed by, group membership of the feedback source. While many situations may conform to this experimental paradigm, there are also many real-life instances in which recipients of feedback are unaware, or uncertain, of the identity of the source of their feedback. This raises the question of whether identity is routinely inferred from the content of feedback, in particular whether feedback providers are seen as ingroup or outgroup members based on what they say.

This reasoning implies that not only categorization of feedback sources frames responses to the content they deliver but also that such feedback content can lead to particular categorization of the source of feedback. Indeed, previous research hints that group members are often motivated to use available cues to re-categorize sources of unflattering feedback and dismiss their comments as a result. For example, negative comments made by ingroup members who have previously demonstrated weak group identification on other indices are treated as if these comments came from outside the group (Hornsey, 2005). However, in these studies participants were still explicitly provided with categorically relevant information prior to receiving criticism. Despite the clues offered by previous research, the effects of feedback content on categorization of anonymous sources remain to be tested directly.

In the research reviewed above, we make the point that (1) the content of feedback (both valence and attributions) may play a key role in the inferences that are made about sources of feedback and (2) that group image threat may represent a central process behind such inferences. In the following section, we consider literature on the role of threat in the categorization of social stimuli more generally before integrating these ideas with the criticism literature and presenting our hypotheses.

Categorization of social stimuli and threat

To simplify the social world and facilitate one's interactions with it, people routinely engage in categorization of social stimuli. Allocation of diverse stimuli into a smaller number of fixed categories achieves the dual-aim of efficient information processing and minimizing cognitive resource expenditure (Macrae & Bodenhausen, 2000). Categorization is also viewed as a fundamental social process (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) – in fact, allocation of a person to a particular social category (i.e., engaging in social categorization) is treated as a basis for a number of social psychological phenomena, such as prejudice and ingroup bias (e.g., Gaertner & Dovidio, 2005).

Notably, the focus of social psychological research has recently been on the consequences of social categorization more than the process of categorization itself (e.g., Rabinovich, Morton, Postmes, & Verplanken, 2012). Self-categorization theory postulates that categorizations are flexible and context-dependent (Turner, Oakes, Haslam, & McGarty, 1994; see also Quinn & Macrae, 2005) and that assigning certain stimuli to 'ingroup' or 'outgroup' depends on both comparative fit (the extent to which these stimuli are perceived as similar to or different from the rest of category members; Turner, 1999;

Turner *et al.*, 1987) and normative fit (the extent to which stimuli fit the stereotype of a particular category; Brown & Turner, 2002). The process of assigning stimuli to social categories is also known to be affected by characteristics of the perceiver (i.e., 'perceiver readiness'), which may be conditioned by individual differences in prejudice (Blascovich, Wyer, Swart, & Kibler, 1997; Hugenberg & Bodenhausen, 2004) or identification with specific groups (Castano, Yzerbyt, Bourguignon, & Seron, 2002).

An additional factor implicated in the process of social categorization is the presence of threat in the environment (cf. Ackerman *et al.*, 2006). Because threat is more likely to come from outgroup members than from inside of one's own group (Baer & McEachron, 1982), threat and 'outgroups' have become reliably associated in certain contexts (e.g., Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001). This would suggest that ambiguous social stimuli associated with threat are likely to have better comparative and normative fit with outgroup (rather than ingroup) categories, and as a result, should be more easily perceived as (typical) outgroup members.

Recent research by Miller, Maner, and Becker (2010) supports this suggestion. In a set of studies, these authors investigated how categorization of racially ambiguous social stimuli (faces or voices) is affected by the degree of threat that they represent. It was demonstrated that increased physical threat (operationalized as increased masculinity, approaching movement, or expression of anger) led to a higher likelihood of targets being perceived as outgroup members (i.e., White participants categorizing racially ambiguous threatening faces and voices as Black). These findings demonstrate that the process of drawing the boundaries between 'us' and 'them' is affected by perceived threat. Although this research specifically focussed on physical threat, it would seem reasonable to assume that other types of threat may have similar implications for social categorization processes. In this article, we consider how group image threats activated by various types of group-directed feedback might determine the social categorization of ambiguous feedback sources.

Present research

In this article we aim to investigate how the content of group-directed feedback affects the categorization of feedback sources as ingroup or outgroup members. Previous research suggests that certain types of feedback (such as internally attributed criticism and externally attributed praise) are associated with group image threat (Rabinovich & Morton, 2010; Rabinovich, Morton, Crook *et al.*, 2012). In addition, research in the domain of social cognition, has demonstrated that ambiguous threatening stimuli are more likely to be perceived as outgroup members (Miller *et al.*, 2010). On the basis of these previous findings, we expect to find a main effect of feedback valence, where critics of one's group will be more readily categorized as outgroup members compared to those who praise the group. In addition, we hypothesize that there will be an interaction between valence of feedback and its attributional content on participants' perception of sources of such feedback as ingroup versus outgroup members. Specifically, *critics* of one's group will be more likely to be perceived as outgroup members when they use internal (group's character) rather than external (group's circumstances) attributions for group's failure. In contrast, sources of *praise* will be more likely to be seen as outgroup members when they use external (rather than internal) attributions for group's success. Finally, consistent with the assumption that threat is the mechanism behind these different categorizations of feedback sources, we expect the combined effects of feedback valence and attributional content to be mediated by group image threat.

We begin by testing these suggestions in the context of feedback provided by an anonymous source on a nation's environmental performance in Study 1.

STUDY 1

Method

Participants and design

Participants were 116 British adults (37 men and 79 women, $M_{\text{age}} = 36.15$, $SD = 15.66$), approached in public places and asked to complete a questionnaire. Participants were randomly assigned to one of the conditions of a 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) between-subjects experimental design. The dependent variable was the categorization of the source of feedback as an ingroup (British) or an outgroup (foreign) member.

Procedure and materials

The study was presented as a survey looking at people's responses to news items. To manipulate feedback type and attribution, participants were presented with a fake newspaper article describing Britain's environmental performance. The article specified that according to a recent review, Britain was either performing well (i.e., praise) or poorly (i.e., criticism) in the environmental domain. In the internal attribution condition, the article went on to explain this performance with reference to the internal characteristics of British people (e.g., 'weak green attitudes' and 'lack of good will' were mentioned in the criticism condition, and 'strong environmental attitudes and values' in the praise condition). In the external attribution condition, Britain's environmental performance was explained with reference to available facilities, and provision of information and incentives (see Rabinovich & Morton, 2010, for similar manipulations).¹ No information about the source of the article or its author was given. After reading the manipulation text, participants were asked four open-ended manipulation check questions to make sure that they had read and understood the article (e.g., 'What was the general feedback on British environmental performance?').

Categorization of the source of feedback was then measured. Participants were asked to estimate how likely it was that the author of the article was 'from the United Kingdom' and 'from outside of the United Kingdom'. Participants responded on a 7-point scale from 1 (*very unlikely*) to 7 (*very likely*). The second item was reverse coded, and the two items were averaged to form a single measure of categorization, where higher values indicate stronger ingroup categorization, $r(114) = 0.84$, $p < .001$.

Results

All participants were able to correctly reproduce the content of the feedback. A 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) ANOVA on the

¹ As noted previously, 'external' attributions refer to incentives or limitations at an institutional (national) level. Although these are clearly not dispositional (i.e., are not referring to ingroup's character, unlike 'internal' attributions), they are also not fully external to participants' ingroup (e.g., are referring to the group's institutional practices). Nonetheless, we retain the labels 'internal' and 'external' for these two types of attributions to maintain consistency with previous research that used very similar manipulations of attributional content (Rabinovich & Morton, 2010; Rabinovich, Morton, Crook et al., 2012).

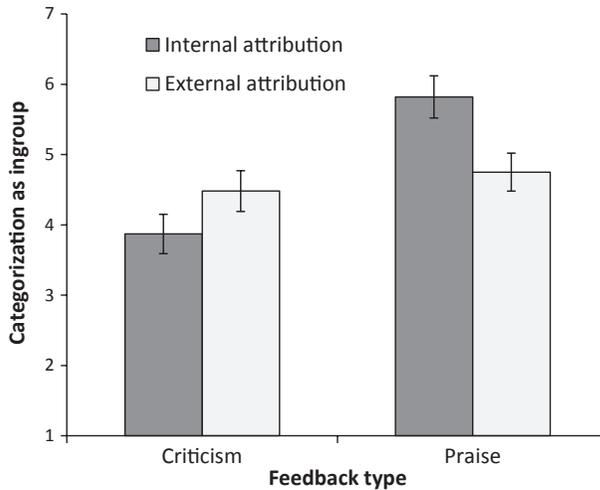


Figure 1. Categorization of feedback source as a function of feedback valence and content (Study 1).

measure of source categorization revealed a significant main effect of feedback type, $F(1, 115) = 14.47, p < .001, \eta^2_p = .11$. The source of praise was more likely to be categorized as an ingroup member ($M = 5.29, SD = 1.61$), than the source of criticism ($M = 4.18, SD = 1.62$).

This was qualified by a significant feedback type by attribution interaction, $F(1, 115) = 8.36, p = .005, \eta^2_p = .07$, see Figure 1. Follow-up comparisons revealed that source of praise was more likely to be categorized as an ingroup member when they used internal ($M = 5.82, SD = 1.61$) rather than external attributions ($M = 4.75, SD = 1.45$), $F(1, 112) = 6.51, p = .012, \eta^2_p = .06$. The source of criticism was more likely to be categorized as an ingroup member when using external ($M = 4.48, SD = 1.64$) rather than internal ($M = 3.87, SD = 1.58$) attributions, although this difference did not reach statistical significance, $F(1, 112) = 2.31, p = .131, \eta^2_p = .02$. Put differently, when internal attributions were made, participants were more likely to categorize the source as an ingroup member when this involved praise rather than criticism, $F(1, 112) = 22.41, p < .001, \eta^2_p = .17$. However, when external attributions were made, the type of feedback did not have a significant effect on categorization, $F(1, 112) = 0.42, p = .520, \eta^2_p < .01$.

Discussion

Study 1 provides initial support for our hypotheses. In particular, it demonstrates that the valence and attributional content of feedback combine to affect categorization of feedback sources. Sources of group-directed criticism are more likely to be categorized as outgroup members as compared to sources of praise. Furthermore, sources of praise were more likely to be seen as outgroup members when they used external (rather than internal) attributions for the group's success, while the opposite pattern was observed for sources of criticism.

² Unless otherwise stated, all comparisons remain statistically significant after the Bonferroni correction for family-wise error is applied.

In Study 1, participants were given no information about the group affiliation of the feedback source. However, a more realistic situation is perhaps one in which feedback recipients have ambiguous, rather than no, information about those who deliver criticism or praise. According to self-categorization theory, *any* person can be categorized along multiple dimensions, and depending on context (i.e., comparative and normative fit) boundaries between ingroups and outgroups can be re-charted (Turner *et al.*, 1994). Therefore, it is reasonable to suggest that not only anonymous threatening sources can be more easily categorized as outgroup members but also sources with ambiguous identity can be *re-categorized* in response to threat. Study 2 explores this possibility. In addition, it explores ingroup threat as a process behind the effects of feedback content.

STUDY 2

Method

Participants and design

Ninety-eight psychology students of a British University took part in the study (82 women, 16 men, $M_{\text{age}} = 22.43$, $SD = 6.64$). Participants were approached on University campus, and randomly assigned to a 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) between-subjects experimental design. The dependent variables were perceived ingroup typicality of the feedback source and perceived ingroup threat.

Procedure and materials

The study was presented as a survey looking at students' perception of information posted on online forums. Participants were presented with a text edited to look like a screenshot from an online forum. The screenshot contained a question from a prospective student (a forum member) who enquired about the examination performance and career prospects of those studying psychology at the participants' university. We manipulated feedback type and attributional content by altering the response to this question provided by another forum member. In all conditions, the person responding ('Alex', a gender-neutral name in the United Kingdom) made it clear that they were studying psychology but in a different UK University (the name of another university situated in the same geographical region and having a similar status was given). Therefore, this person could be seen by the participants as an ingroup member (another psychology student) or as an outgroup member (a student from a different university). The commenter also made it clear that they had some experience with the psychology course at the participants' university through studying there for a term on an exchange.

In the negative feedback condition, the commenter said that his/her overall impression was not favourable. It was mentioned that examination results were generally poor, and psychology students at the participants' university seemed to have problems with finding jobs after graduation. In the positive feedback condition, the opposite comment was made – the commenter mentioned great examination results and excellent graduate prospects. In the internal attribution condition, the commenter went on to explain these failures or success by internal characteristics of psychology students at this university (e.g., students' high abilities and hard work or lack of these). In the external attribution condition, the same outcomes were explained by available facilities and the amount of contact hours. After reading the manipulation text, participants were asked four open-ended manipulation check questions to make sure that they had read and

understood the text. These included the questions that required participants to reproduce the commenter's place and subject of study and the essence of their comment.

Following this, perceived ingroup typicality and ingroup threat were measured. Participants responded to all items on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). One item was used to measure ingroup typicality: 'Alex is a typical psychology student'. Psychology was chosen as the relevant category because this is the group membership that participants shared with the feedback source.

To measure perceived ingroup threat, four items were used: for example, 'I feel that the image of psychology students in [participants' university] is under threat', 'I feel that the image of psychology students in [participants' university] is in danger', $\alpha = .87$. After completing the questionnaire, participants were thanked and debriefed.

Results

All participants were able to correctly reproduce the content of feedback and affiliation of the feedback source.

A 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) ANOVA on perceived ingroup typicality of the source showed a significant main effect of feedback type, $F(1, 98) = 39.32, p < .001, \eta^2_p = .30$: The feedback source was seen as a more typical ingroup member when they provided praise ($M = 4.82, SD = 0.80$) rather than criticism ($M = 3.58, SD = 1.14$). This main effect was qualified by a significant feedback type \times attribution interaction, $F(1, 98) = 5.42, p = .022, \eta^2_p = .05$, see Figure 2. Follow-up comparisons showed that the source of criticism was seen as a slightly more typical ingroup member when they used external ($M = 3.87, SD = 1.18$) rather than internal ($M = 3.32, SD = 1.07$) attributions, $F(1, 98) = 3.87, p = .052, \eta^2_p = .04$. At the same time, the source of praise was seen as a more typical ingroup member when they used internal ($M = 5.00, SD = 0.76$) rather than external ($M = 4.64, SD = 0.81$) attributions, but this difference did not reach statistical significance, $F(1, 98) = 1.73, p = .191, \eta^2_p = .02$. Put differently, although participants always perceived the feedback source as a more typical ingroup member when they provided praise rather than criticism, this effect was stronger when internal, $F(1, 98) = 37.77, p < .001, \eta^2_p = .29$, versus external attributions were used, $F(1, 98) = 7.61, p = .007, \eta^2_p = .08$.

The same ANOVA was conducted on the measure of perceived ingroup threat. There was a significant main effect of feedback type, $F(1, 98) = 37.56, p < .001, \eta^2_p = .29$: Participants were more threatened by criticism ($M = 4.49, SD = 1.14$) than praise ($M = 2.76, SD = 1.75$). Again, this main effect was qualified by a significant interaction, $F(1, 98) = 10.45, p = .002, \eta^2_p = .10$, see Figure 3. Follow-up pairwise comparisons showed that participants in the praise condition reported a higher level of ingroup threat when praise was attributed externally ($M = 3.46, SD = 1.69$) rather than internally ($M = 2.03, SD = 1.53$), $F(1, 98) = 12.79, p = .001, \eta^2_p = .12$. When criticism was instead given, attributions did not have a significant effect on perceived ingroup threat: $M_{\text{internal}} = 4.69, SD = 1.03$; $M_{\text{external}} = 4.28, SD = 1.24$; $F(1, 98) = 1.02, p = .316, \eta^2_p = .01$. Put differently, participants always perceived criticism as more threatening than praise, but this effect was stronger when internal, $F(1, 98) = 44.30, p < .001, \eta^2_p = .32$, versus external attributions were used, $F(1, 98) = 4.15, p = .045, \eta^2_p = .04$.

³ This comparison becomes non-significant after applying the Bonferroni correction for family-wise error.

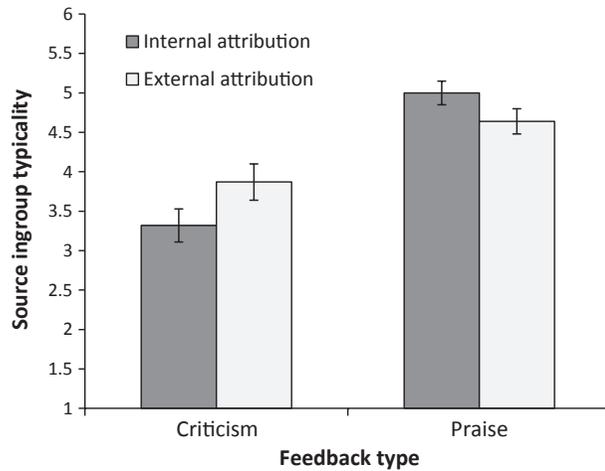


Figure 2. Ingroup typicality of feedback source as a function of feedback valence and content (Study 2).

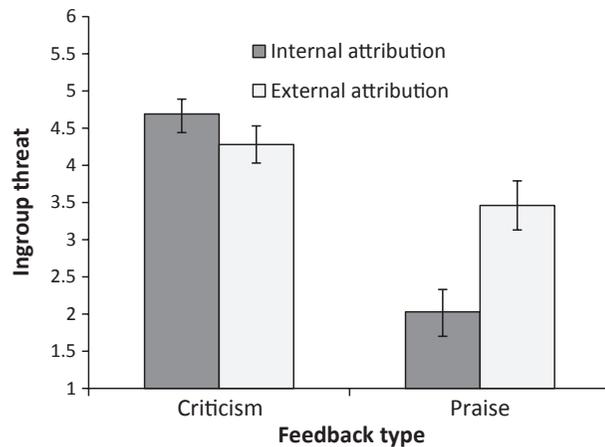


Figure 3. Ingroup threat as a function of feedback valence and content (Study 2).

Mediation

The above analyses demonstrated similar effects of feedback type and attributional content on perceived ingroup threat and perceived ingroup typicality of the feedback source. Therefore, it was possible that perceived threat mediated the effects of the independent variables on ingroup typicality (i.e., the most threatening sources of feedback are categorized as the least typical ingroup members). To explore this possibility, we conducted mediated moderation analysis (Muller, Judd, & Yzerbyt, 2005). First, we regressed the ingroup typicality (the dependent variable) on attributional content (the independent variable), feedback type (the moderator), and their interaction. Second, we regressed perceived ingroup threat (the mediator) on the same predictors. Finally, we regressed ingroup typicality on feedback type, attributional content, their interaction, and perceived ingroup threat (centred).

The results met the conditions for mediated moderation. Specifically, the interaction between feedback type and attribution was a significant predictor of both the ingroup typicality of the source, $\beta = .334$, $p = .022$, and ingroup threat, $\beta = .458$, $p = .002$. The effect of perceived threat on ingroup typicality of the source was significant after controlling for the other predictors, $\beta = -.288$, $p = .006$, and the effect of the feedback type by attribution interaction was reduced with the mediator included in the equation, $\beta = .202$, $p = .174$. To further establish the case for mediation, we conducted a bootstrapping analysis using the PROCESS macro (Hayes, 2012). The bias-corrected bootstrap estimate of the indirect effect of the interaction between feedback type and attribution had a 95% confidence interval of 0.0618 to 0.7237. The mediation was significant in the praise condition ($[-0.5909; -0.0587]$), but not in the criticism condition ($[-0.0502; 0.2571]$): Within the praise condition, internal attribution had an indirect positive effect on ingroup categorization via decreased threat, whereas in the criticism condition, threat did not mediate the relationship between attribution and categorization.

Discussion

Study 2 replicated the results of Study 1 in a situation where participants received ambiguous information about the source of feedback rather than no information at all. As in Study 1, a significant interaction between feedback valence and attributional content was observed on categorization of the feedback source as a typical ingroup member (i.e., psychology student). Sources of praise were categorized as more typical ingroup members when they used internal (as opposed to external) attributions for success. However, we did not observe a significant effect of attributional content on perceptions of critics, although the pattern of means was in the predicted direction. Importantly, and in line with our predictions, Study 2 demonstrated that the effect of feedback content on categorization of a feedback source was mediated by perceived group image threat: Less threatening sources were more likely to be categorized as typical ingroup members.

A limitation of Study 2 is that it is possible to point at a confound of the attributional content manipulation: In the internal attribution condition, the explanation of failure or success referred to the category that participants shared with the commenter (psychology students, albeit in a different university), while in the external attribution condition the explanation referred to the entity that was not shared (participants' university). Study 3 aims to compensate for this limitation by using a cleaner manipulation of attributional content. In addition, it includes a measure of outgroup typicality of the source to supplement the previous findings on perceived ingroup typicality.

STUDY 3

Method

Participants and design

One hundred and fifty-eight first-year psychology students of a British University took part in the study (130 female, 28 male, $M_{\text{age}} = 18.83$, $SD = 2.52$). All participants were British. Participants took part in the study during a practical class. They were randomly assigned to a 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) between-subjects experimental design. The dependent variables were perceived ingroup and outgroup typicality of the feedback source, and perceived ingroup threat.

Procedure and materials

The study was presented in the same way as Study 2, and the same manipulations of feedback valence and attributional content were used. There were two main differences, however: First, the forum question and the feedback provided in response to it concerned *students at the participants' University* (not *psychology* students at this university, as in Study 2); and second, the source of feedback was presented as an international student from Armenia studying in the participants' university. This person could be seen by our British participants as an ingroup member (another student at the same university) or as an outgroup member (a foreigner). Crucially, unlike in Study 2, explanations of failure or success in both internal and external attribution conditions referred to the group that the commenter belonged to (students of a particular university or the university itself). After reading the manipulation, participants were asked a number of open-ended manipulation check questions, where they were required (among other things) to identify the commenter's nationality and place of study, and to summarize their comment.

Following this, perceived ingroup and outgroup typicality, and ingroup threat were measured. Participants responded to all items on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). One item was used to measure ingroup typicality: 'Alex is a typical student of participants' University' another item measured outgroup typicality: 'Alex is a typical Armenian'. These two items were moderately positively correlated, $r(156) = .25$, $p = .002$, and consequently were analysed as two separate measures rather than being collapsed into a single index of ingroup versus outgroup categorization. Finally, perceived ingroup threat was measured in the same way as in Study 2 with the items rephrased so that they referred to 'students of participants' University' ($\alpha = .90$). After completing the questionnaire, participants were thanked and debriefed.

Results

All participants were able to correctly reproduce the content of feedback and nationality of the feedback source. To test the effect of feedback type and attributional content on perceived ingroup and outgroup typicality of the source, we computed a difference score to represent the relative categorization of the source as ingroup versus outgroup by subtracting outgroup typicality score from ingroup typicality score. Positive scores on this measure reflect categorization of the source as more ingroup than outgroup, whereas negative scores indicate categorization of the source as more outgroup than ingroup. A 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) ANOVA on this difference score revealed a main effect of feedback, $F(1, 157) = 57.54$, $p < .001$, $\eta^2_p = .27$, indicating that sources of praise were perceived as more typically ingroup members ($M = 1.17$, $SD = 1.74$), while sources of criticism were seen as more typically outgroup members ($M = -0.79$, $SD = 1.53$). This effect was qualified by a significant feedback valence \times attribution interaction, $F(1, 157) = 9.51$, $p = .002$, $\eta^2_p = .06$, indicating that the main effect of feedback valence was stronger in the internal, $M_{\text{praise}} = 1.60$, $SD = 1.64$; $M_{\text{criticism}} = -1.10$, $SD = .51$; $F(1, 157) = 59.93$, $p < .001$, $\eta^2_p = .28$, rather than external attribution condition, $M_{\text{praise}} = 0.65$, $SD = 1.74$; $M_{\text{criticism}} = -0.49$, $SD = 1.50$; $F(1, 157) = 9.65$, $p = .002$, $\eta^2_p = .06$ (see Figure 4). Put differently, sources of praise were always seen as more typical ingroup (rather than outgroup) members, but more so when they used internal rather than external attribution, $F(1, 157) = 6.66$, $p = .011$, $\eta^2_p = .04$. Similarly, critics were always seen as more typical outgroup (rather than ingroup) members, but marginally more so when they used internal rather than external attributions, $F(1, 157) = 3.09$, $p = .081$, $\eta^2_p = .02$.

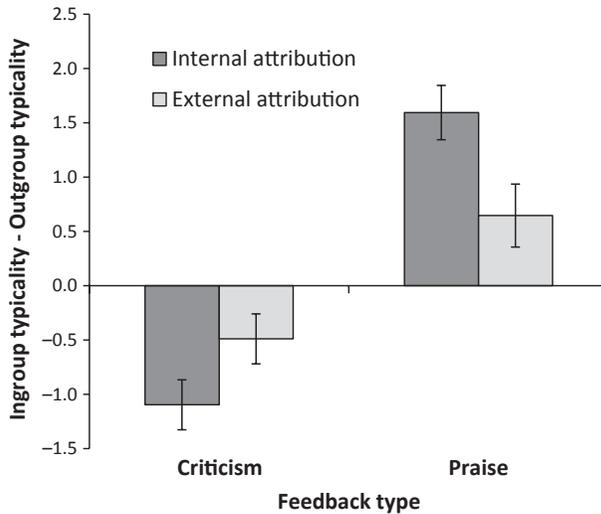


Figure 4. Difference between ingroup and outgroup typicality as a function of feedback valence and content (Study 3).

The same 2 (feedback type: criticism vs. praise) \times 2 (attribution: internal vs. external) ANOVA was conducted on the measure of threat. The analysis revealed a significant main effect of feedback valence, whereby participants reported stronger perception of threat in response to criticism rather than praise, $M_{\text{criticism}} = 4.41$, $SD = 1.29$; $M_{\text{praise}} = 2.13$, $SD = 1.26$; $F(1, 157) = 155.26$, $p < .001$, $\eta^2_p = .50$, and a significant main effect of attributional content whereby participants perceived more threat when external rather than internal attributions were used, $M_{\text{internal}} = 3.06$, $SD = 1.95$; $M_{\text{external}} = 3.62$, $SD = 1.36$; $F(1, 157) = 6.76$, $p = .010$, $\eta^2_p = .04$. These main effects were, however, qualified by a significant feedback valence \times attributional content interaction, $F(1, 157) = 48.41$, $p < .001$, $\eta^2_p = .24$, see Figure 5. Follow-up pairwise comparisons showed that participants in the praise condition reported a higher level of ingroup threat when praise was attributed externally ($M = 3.07$, $SD = 1.31$) rather than internally ($M = 1.40$, $SD = 0.51$), $F(1, 157) = 43.27$, $p < .001$, $\eta^2_p = .22$. In the criticism condition, participants perceived more threat when internal ($M = 4.80$, $SD = 1.23$) rather than external ($M = 4.04$, $SD = 1.25$) attributions were used, $F(1, 157) = 10.05$, $p = .002$, $\eta^2_p = .06$. Put differently, participants always perceived criticism as more threatening than praise, but this effect was stronger when internal, $F(1, 157) = 197.41$, $p < .001$, $\eta^2_p = .56$, rather than external, $F(1, 157) = 14.49$, $p < .001$, $\eta^2_p = .08$, attributions were used.

Mediation

To explore the prediction that perceived threat mediates the effect of feedback valence and attribution on the difference between ingroup and outgroup typicality of the feedback source, we conducted the same mediated moderation analysis as in Study 2 with the difference score between ingroup and outgroup typicality as a dependent variable. The results again met the conditions for mediated moderation. Specifically, the interaction between feedback type and attribution was a significant predictor of both the difference between ingroup and outgroup typicality, $\beta = .361$, $p = .002$, and ingroup threat, $\beta = .624$, $p < .001$. The effect of perceived threat on the difference between ingroup and

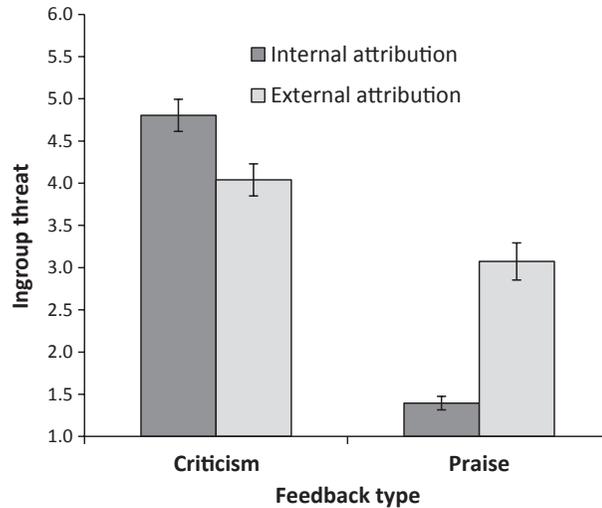


Figure 5. Ingroup threat as a function of feedback valence and content (Study 3).

outgroup typicality was significant after controlling for the other predictors, $\beta = .265$, $p = .010$, and the effect of the feedback type by attribution interaction was reduced with the mediator included in the equation, $\beta = .188$, $p = .155$. The bias-corrected bootstrap estimate of the indirect effect of the interaction between feedback type and attribution had a 95% confidence interval of -1.4216 to -0.2042 . The mediation was significant in both criticism ($[-0.5870; -0.0407]$) and praise ($[0.1380; 0.9428]$) conditions.

Discussion

Study 3 replicated the results on Study 2 with a cleaner manipulation of attributional content, and a dependent measure that captures both perceived ingroup and outgroup typicality of the feedback source. In line with the predictions, it demonstrated that sources of praise are perceived as more typical ingroup rather than outgroup members when they use internal (rather than external) attributions for success, while the opposite is true for critics. In addition, Study 3 replicated the mediation via perceived ingroup threat. We consider the implications of these findings in the General Discussion.

GENERAL DISCUSSION

This study aimed to explore the effects of feedback content on the social categorization of feedback sources. While previous research on responses to feedback has focused on situations in which information about the social affiliation of feedback sources is explicitly provided, many real-life contexts differ from this set up: Sources of feedback are often either not known or ambiguous in terms of their affiliations. In these contexts, rather than being a primary parameter that determines the interpretation of feedback, source affiliation (i.e., categorization) may be a secondary inference made on the basis of the content of their feedback. In three experimental studies, we tested whether such inferences depend systematically on the valence and attributional content of feedback.

The main finding of this research is that audiences do draw different conclusions about the group membership of a feedback source depending on the type of feedback that they

provide. In particular, critics are more likely to be perceived as outgroup members than those who provide praise. In addition to this basic effect, the attributions that feedback providers make when explaining the group's success or failure also play an important role in determining how they are perceived. The reliable interplay between feedback valence and attributional content suggests that sources of feedback are more likely to be perceived as insiders when they use internal (rather than external) attributions for success, whereas the opposite seems true for sources of criticism. Importantly, consistent with our reasoning, Studies 2 and 3 demonstrated that the above effects are mediated through perceptions of group image threat: Sources of more threatening feedback (i.e., internally attributed criticism and externally attributed praise) were more readily categorized as outgroup members.

Implications

The present findings extend previous research on responses to group-based feedback. In particular, they shed some light on psychological processes that occur in the highly realistic situation where feedback is given without a specific affiliation of the commenter being made salient from the outset. Our findings demonstrate that when such information is not available, it may be inferred from the content of feedback. The present findings are consistent with previous research on the role of attributional content of feedback. Once again, they demonstrate that the attributions given for success or failure can considerably alter perceptions of otherwise identical feedback (Rabinovich & Morton, 2010, Rabinovich, Morton, Crook *et al.*, 2012). Taken together, previous research on responses to feedback and the present findings provide a comprehensive picture of the processes unfolding in response to group-based feedback, including inferences about sources of feedback (present research), and the impact of these inferences on feedback reception (previous research).

The present findings corroborate recent research showing that more threatening targets are less likely to be categorized as ingroup members (Miller *et al.*, 2010) and extend these findings by demonstrating that the categorization processes in question can be extrapolated beyond physical threat – specifically, to group image threat. This study provides an important link between research on social categorization and research on responses to group-based feedback by demonstrating how certain types of feedback *create* threat, and how this sense of threat translates into specific inferences about the group membership of the feedback source.

At this point, it may be interesting to consider how the categorization processes outlined in this research relate to the strategic processes described in previous research (e.g., expressing and enacting a stronger need for reform in response to internally attributed outgroup criticism, Rabinovich & Morton, 2010). It is possible that the two processes operate in sequence: Re-categorizing threatening sources of feedback as outgroup members can (paradoxically) lead recipients of feedback to consider their criticism or praise more carefully. However, it would also seem reasonable to suggest that not all outgroups possess sufficient power and authority to motivate ingroup members to respond to their feedback in a strategic way (i.e., by demonstrating improved performance; see Klein, Spears, & Reicher, 2007, on strategic responding to various types of outgroup audiences). Instead, it is likely that recipients of feedback who engage in re-categorization of threatening sources would choose to see them as members of outgroups with least authority. In this case, re-categorization of a feedback source as an outgroup member is unlikely to be followed by a strategic improvement in performance in

response to their feedback. Thus, power may further moderate whether people become responsive to, versus dismissive of, threatening outgroup feedback.

The main practical implication of the present findings is that those who provide negative feedback to groups face a tough task. It seems that even when criticism is delivered by an insider, group members may find ways to re-categorize the source of such feedback as an outgroup member. Our findings corroborate the previous research that demonstrates that delivering group-based criticism is risky and may result in sources of criticism being ostracized (cf. Hornsey, 2005). At the same time, they suggest that this risk can be accentuated or attenuated depending on attributions used.

Limitations and further research

One of the limitations of this research is that it stops short of fully exploring consequences of feedback source categorization. Re-categorization of feedback sources may have consequences not only for the way the source itself is perceived but also for self-perception of feedback targets. For example, receiving negative feedback from an ambiguous source may decrease one's identification with a category that is *shared* with the source of feedback, while increasing identification with a category that is not shared (e.g., participants in the criticism condition of Study 3 could have felt increased identification with their nation and decreased identification with their university). In addition, receiving feedback from an ambiguous source may affect the way one's own group is perceived and alter the criteria for ingroup membership. For example, if a British person from an ethnic minority criticizes Britain, could this result in the feedback recipients' excluding ethnic minorities from the category 'British' (e.g., construing 'the British' as essentially White category)? Along these lines, future research could explore whether re-categorization of individual group members has wider consequences for mapping the borders between the groups.

Although this study demonstrates that the data consistently support the idea that the effect of feedback content on source categorization is mediated by ingroup threat, the question remains why exactly threatening sources tend to be categorized as outgroup members. Previous theorizing seems to suggest that this tendency is developed through repeated experience with ingroup and outgroup members in the process of which people learn that outgroups are more likely to represent threat (Miller *et al.*, 2010). There is, however, a possibility that the categorization process is more strategic – perhaps, sources of threat can be psychologically distanced and neutralized by categorizing them as outgroup members. The precise process behind the link between threat and outgroup categorization remains to be investigated.

In addition, while the data are consistent with the suggestion that threat mediates the effect of feedback content on categorization of the feedback source, they are also not inconsistent with an alternative suggestion that categorization mediates the effect on threat: Some sources of feedback may be seen as more threatening as a result of being categorized as outgroup members. Future research could clarify the direction of the mediation by manipulating threat or categorization directly. It is also possible that the influence between the two processes is reciprocal: Increased threat leads to outgroup categorization, which leads to further increase in perceived threat.

Another limitation of the present findings is that it may not be applicable to groups with strong self-critical norms. Groups that encourage expressions of critical feedback would be unlikely to re-categorize its critical members as outsiders. Indeed, where criticism is compatible with group norms, expressing it may prove a path towards ingroup

acceptance (see Postmes, Spears, & Cihangir, 2001). Future research could explore the effect of group norms around the acceptability of criticism on the observed effects.

Conclusion

Previous research has demonstrated that the group membership of a person providing group-directed feedback has a significant impact on how this feedback is received. However, in many real-world situations sources of feedback are members of multiple social groups and can be categorized by feedback recipients as ingroup members on some dimensions, but as outgroup members on the others. This study demonstrates that the categorization of feedback sources is guided by the valence and attributional content of their feedback. Specifically, critics are more readily seen as outgroup members than those who provide praise. In addition, sources of praise are more likely to be categorized as ingroup members when they attribute success to the group's internal qualities rather than their external circumstances, while the opposite trend seems true for sources of criticism. Importantly, the interactive effect of feedback valence and attributional content is mediated through perceived group image threat. The sources of feedback that threaten the group's image most strongly are the ones that are most likely to be categorized as outgroup members. These findings demonstrate that information about the group membership of feedback sources does not have to precede feedback (as it does in the paradigm used in previous research on group-directed feedback), but that this can also be an outcome that is inferred from the content of feedback. This would suggest that when group membership of the source is ambiguous, feedback content may be the principal determinant of responses to feedback.

Acknowledgement

We thank Matthew Hornsey for the very helpful discussion of the ideas included in this article.

References

- Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., ... Schaller, M. (2006). They all look the same to me (unless they're angry). *Psychological Science*, *17*, 836–840. doi:10.1111/j.1467-9280.2006.01790.x
- Baer, D., & McEachron, D. L. (1982). A review of selected sociobiological principles: Application to hominid evolution I: The development of group structure. *Journal of Social & Biological Structures*, *5*, 69–90. doi:10.1016/S0140-1750(82)91588-3
- Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, *80*, 253–267. doi:10.1037//0022-3514.80.2.253
- Blascovich, J., Wyer, N. A., Swart, L. A., & Kibler, J. L. (1997). Racism and social categorization. *Journal of Personality and Social Psychology*, *72*, 1364–1372. doi:10.1037/0022-3514.72.6.1364
- Brown, P. M., & Turner, J. C. (2002). The role of theories in the formation of stereotype content. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (pp. 67–89). Cambridge, UK: Cambridge University Press.
- Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology*, *38*, 315–322. doi:10.1006/jesp.2001.1512

- Gaertner, S. L., & Dovidio, J. F. (2005). Categorization, recategorization and intergroup bias. In J. F. Dovidio, P. Glick, & L. Rudman (Eds.), *Reflecting on the nature of prejudice* (pp. 71–88). Philadelphia, PA: Psychology Press.
- Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable moderation, mediation, and conditional process modeling*. Retrieved from <http://afhayes.com/spss-sas-and-mplus-macros-and-code.html>
- Hornsey, M. J. (2005). Why being right is not enough: Predicting defensiveness in the face of group criticism. *European Review of Social Psychology*, *16*, 301–334. doi:10.1080/10463280500436040
- Hornsey, M. J., & Imani, A. (2004). Criticising groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect. *Personality and Social Psychology Bulletin*, *30*, 365–383. doi:10.1177/0146167203261295
- Hornsey, M. J., Trembath, M., & Gunthorpe, S. (2004). “You can criticise because you care”: Identity attachment, constructiveness, and the intergroup sensitivity effect. *European Journal of Social Psychology*, *34*, 499–518. doi:10.1002/ejsp.212
- Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342–345.
- Klein, O., Spears, R., & Reicher, S. (2007). Social identity performance: Extending the strategic side of SIDE. *Personality and Social Psychology Review*, *11*, 28–45. doi:10.1177/1088868306294588
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. doi:10.1146/annurev.psych.51.1.93
- Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-Protective biases in group categorization: Threat cues shape the psychological boundary between “us” and “them.” *Journal of Personality and Social Psychology*, *99*, 62–77. doi:10.1037/a0018086
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863. doi:10.1037/0022-3514.89.6.852
- Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology*, *80*, 918–930. doi:10.1037/0022-3514.80.6.918
- Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of personal construal. *Journal of Personality and Social Psychology*, *88*, 467–479. doi:10.1037/0022-3514.88.3.467
- Rabinovich, A., & Morton, T. A. (2010). Who says we are bad people? The impact of criticism source and attributional content on responses to group-based criticism. *Personality and Social Psychology Bulletin*, *36*, 524–536. doi:10.1177/0146167210362980
- Rabinovich, A., & Morton, T. A. (2012). Sizing fish and ponds: The joint effects of individual- and group-based feedback. *Journal of Experimental Social Psychology*, *48*, 244–249. doi:10.1016/j.jesp.2011.07.017
- Rabinovich, A., Morton, T. A., Crook, M., & Travers, C. (2012). Let another praise you? The effects of source and attributional content on responses to group-directed praise. *British Journal of Social Psychology*, *51*, 753–761. doi:10.1111/j.2044-8309.2011.02093.x
- Rabinovich, A., Morton, T. A., Postmes, T., & Verplanken, B. (2012). Collective self and individual choice: The effects of intergroup comparative context on environmental values and behaviour. *British Journal of Social Psychology*, *51*, 551–569. doi:10.1111/j.2044-8309.2011.02022.x
- Turner, J. C. (1999). Current issues in research on social identity and self-categorization theories. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 6–34). Oxford, UK: Blackwell.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group*. Oxford, UK: Blackwell.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, *20*, 454–463. doi:10.1177/0146167294205002