

# Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance

Paul Hutchison, Dominic Abrams, Roberto Gutierrez, G. Tendayi Viki

## Abstract

Two studies investigated the impact of the presentation of an undesirable group member on group stereotype judgments among participants with varying degrees of ingroup identification. In Study 1 ( $N = 67$ ), identification was associated with stereotype change following presentation of an undesirable, but not a desirable, ingroup member. This relationship was mediated by evaluations of the undesirable ingroup member: the stronger the identification, the more negative the evaluation, and the greater the shift towards a more positive ingroup stereotype. In Study 2 ( $N = 180$ ), identification was positively associated with ingroup stereotype ratings following presentation of an undesirable ingroup member but was negatively associated with outgroup ratings following presentation of an undesirable outgroup member. As in Study 1, the association between ingroup identification and ingroup stereotype ratings was mediated by evaluations of the undesirable ingroup member. Results are discussed in relation to the black sheep effect and identity maintenance strategies.

Keywords: Deviance; Black sheep effect; Stereotype change; Social identity; Group processes; Exclusion

The tendency to favor ingroup members over outgroup members is a pervasive but not necessarily inevitable feature of intergroup relations (e.g., [Brewer, 1979](#) and [Mullen et al., 1992](#)). In some situations ingroup members display the opposite and favor outgroup members over ingroup members. Research on the 'black sheep' effect ([Marques & Yzerbyt, 1988](#)) has showed that desirable ingroup members are favored over similar outgroup members, whereas undesirable outgroup members are favored over similar ingroup members. It has been suggested that devaluation might serve to exclude undesirable members from the representation of the ingroup and protect the group's identity (e.g., [Marques & Paez, 1994](#)). Although this prediction has received considerable circumstantial support from research comparing evaluations of desirable and undesirable ingroup and outgroup members, the link between devaluation and identity maintenance has not yet been established empirically. The current research aimed to address this limitation. In two experiments we investigated the impact of the presentation of desirable and undesirable group members on perceptions of the group among participants who differed in their level of identification with the ingroup.

### *Social identity maintenance*

According to social identity theory ([Tajfel & Turner, 1986](#)), people's identities are derived largely from their memberships in different social categories or groups. When a group membership is salient, individuals evaluate themselves less in terms of their unique and idiosyncratic attributes and more in terms of the comparative properties of the group. To this extent, people are thought to have a basic motivation to see their own groups as superior to outgroups on relevant dimensions. Excluding undesirable members from the ingroup thus serves the important function of maintaining a positive and distinctive social identity ([Marques & Paez, 1994](#)).

Support for these ideas comes from several studies showing that desirable ingroup members are evaluated more favorably than identical outgroup members, whereas undesirable ingroup members are evaluated less favorably than identical outgroup members—the 'black sheep' effect (see [Marques & Paez,](#)

[1994](#), for a review). [Marques and Paez \(1994, p. 38\)](#) suggested that devaluation might serve to exclude undesirable members from the representation of the ingroup (see also [Eidelman, Silvia, & Beirnet, 2006](#)). In this view, the undesirable target is portrayed in such an extremely negative light that he/she can no longer be seen as a typical ingroup member. A presumed consequence of this psychological exclusion is that the overall perception of the ingroup should be enhanced. This follows from research showing contrast and assimilation effects on group-based judgments following exposure to salient group members. If information about salient group members is included in the representation that is formed when judging the group, the resulting perception of the group will assimilate to that information (e.g., [Bless & Schwarz, 1998](#)). Conversely, excluded information produces no assimilation and may be used as a standard against which the group as a whole is judged. Several studies have showed that excluding deviants from the representation of the group has a contrast effect on subsequent group stereotype judgments (e.g., [Kunda & Oleson, 1997](#)). Excluding an undesirable member from the ingroup should therefore enhance the perception of the ingroup's positive stereotypical attributes.

Consistent with this reasoning, [Hutchison and Abrams \(2003\)](#) found that high identifiers expressed a more positive ingroup stereotype after, compared to before, reading about an undesirable ingroup member. High identifiers who read about an undesirable ingroup member also expressed a more positive ingroup stereotype than those presented with a desirable ingroup member, whereas low identifiers' perceptions of the ingroup were unaffected by the target group members. This suggests that high identifiers may have excluded the undesirable member from their representation of the ingroup. This interpretation is supported further by target evaluation ratings, which showed that high identifiers were more negative than low identifiers towards the undesirable ingroup member (see also [Castano, Paladino, Coull, & Yzerbyt, 2002](#)).

*The relationship between group identification, deviant derogation, and identity maintenance*

Although these studies suggest that high identifiers may respond to the presence of an undesirable ingroup member with a motivation to protect their group's identity, evidence for the inferred psychological exclusion process remains circumstantial and indirect. Moreover, the findings from which exclusion is inferred are open to alternative explanation. Group members may derogate an undesirable target extremely but still consider him/her to be a member of the ingroup. Indeed, high identifiers in [Hutchison and Abrams's \(2003\)](#) study may have expressed more dislike for the deviant not necessarily in an attempt to exclude him/her from their representation of the ingroup, but rather because they were more threatened by the undesirable actions of a fellow ingroup member. This threat may also have led high identifiers to enhance their perception of the ingroup's positive stereotypical attributes by way of compensation. To this extent, the inferred link between devaluation and psychological exclusion remains to be established empirically. The current research aimed to address this limitation.

In two studies we assessed the impact of the presentation of an undesirable group member on group stereotype judgments among participants with varying degrees of identification with the ingroup. If devaluation serves to exclude undesirable members from the representation of the ingroup, as traditional explanations of the black sheep effect have argued (e.g., [Marques & Paez, 1994](#)), we would expect the level of identification with the ingroup to impact on stereotype change following presentation of an undesirable ingroup member through evaluations of the undesirable ingroup member. If on the other hand devaluation and stereotype enhancement are unrelated to psychological exclusion, then there is no reason to expect target evaluation ratings to mediate the effect of identification on subsequent ratings of the ingroup. In Study 2, we also examined how an undesirable outgroup member might impact on perceptions of the outgroup stereotype. We reasoned that highly identified participants would be more motivated to exclude an undesirable target from the representation they formed when judging the ingroup than from their representation of the outgroup. This difference, we reasoned, should be reflected in ingroup and outgroup stereotype judgments following presentation of an undesirable ingroup or outgroup member, respectively.

## Study 1

### Method

#### Participants

Sixty-seven psychology students from the University of Kent participated as a course requirement. Fifty-five were female and 12 were male.

#### Materials and procedure

The study was conducted as part of a research methods lecture. After a brief introduction to the lecture, participants completed four items assessing their level of identification with the group 'psychology students' (from [Doosje, Ellemers, & Spears, 1995](#)): "I see myself as a psychology student," "I identify with psychology students as a group," "I am pleased to be a psychology student," and "I feel strong ties with psychology students." Responses to these and subsequent items were recorded on rating-scales ranging from 1 (*not at all*) to 9 (*very much*). The items were combined (averaged) to form a single *group identification* score (Cronbach's  $\alpha = .89$ ).

Participants then rated psychologists as a group on a range of positive and negative stereotypical dimensions (from [Coull,](#)

[Yzerbyt, Castano, Paladino, & Leemans, 2001](#)): "sensitive," "irresponsible," "empathic," "capable of understanding other people's personalities," "unethical," "can be trusted," "unfriendly." The negative items were reverse scored and combined with the positive items to form a single *group stereotype* score ( $\alpha = .93$ ). A higher score indicates a more positive stereotype. The questionnaire also included six filler items on topics such as 'career aims' and 'university life' which are not included in the analyses.

Upon completion, the questionnaires were collected by the researcher and participants were informed that the true purpose of the questionnaire was to familiarize them with rating-scales to allow them to participate in a further study at the end of the lecture. Participants were informed that the purpose of the 'second' study was to investigate people's ability to form impressions of individuals on the basis of varying amounts of information. Participants read one of two descriptions of a target psychologist (from [Coull et al., 2001](#)). Both descriptions began with the same demographic information. The *desirable* target description continued as follows:

"He very carefully listens to his patients in order to fully understand their problems. He finds the right words to help patients to understand the issues, and expresses a lot of warmth and empathy when needed. His strong analytical and synthetic skills help him to elaborate solutions and treatments."

In contrast, the *undesirable* target description continued as follows:

"He tends to see his own problems in the patients' lives. He often interrupts patients because he is nervous, and fails to fully understand their point of view. He lacks the human warmth needed to gain the patients' trust. Finally, he often mixes up patients' records and asks people to explain their problems again and again."

Participants then evaluated the target psychologist using three items: "This person is a good psychologist," "I would like to work with this psychologist," "I would trust this psychologist." These items were combined to form a single *target evaluation* score ( $\alpha = .95$ ). Next, participants re-completed the group stereotype items ( $\alpha = .93$ ) and were asked to write down any suspicions they might have about the true purpose of the study. No accurate suspicions were recorded.<sup>1</sup>

## Results

#### Group stereotype ratings

Our first hypothesis concerned the stereotype ratings. We predicted that ingroup identification would be associated with change towards a more positive ingroup stereotype following presentation of an undesirable ingroup member. We tested this prediction using hierarchical regression analysis. In four steps, we regressed the stereotype ratings recorded after the presentation of the target group members on (1) the pre-presentation stereotype ratings, (2) the standardized identification scores, the target variable (effect-coded  $-1 = desirable, 1 = undesirable$ ), (3) the cross products of these variables, and (4) the three-way interaction term.<sup>2</sup> In what follows, the group stereotype ratings are the post-presentation ratings with the pre-presentation ratings controlled for.

The pre-presentation stereotype main effect ( $\beta = .79, t[60] = 11.45, p < .001$ ) and the Identification  $\times$  Target interaction ( $\beta = .12, t[60] = 2.03, p = .047$ ) were significant, all others  $< 0.77, ps > .45$ . The interaction is represented graphically in [Fig. 1](#), following procedures outlined by [Aitken](#)

and West (1991). To test our specific predictions we compared the slopes of the regression lines for identification in both target conditions. Identification was positively associated with stereotype ratings following presentation of an undesirable ingroup member ( $\beta = .24, t[33] = 2.87, p = .01, R^2 = .86$ ), whereas there was no association between these variables following presentation of a desirable ingroup member ( $\beta = .05, t[30] = .53, p = .60, R^2 = .83$ ). Confirming our prediction, the more participants identified with the ingroup, the more positive they rated the ingroup stereotype after, compared to before, presentation of an undesirable ingroup member.

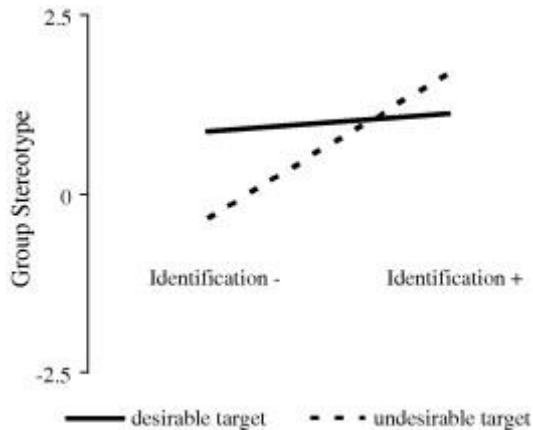


Fig. 1. Graphical representation of interaction between identification and target on group stereotype ratings. *Note.* Group stereotype ratings are the post-presentation ratings with the pre-presentation ratings controlled for. A higher score indicates a more positive stereotype.

#### Target evaluation ratings

To test our second prediction, that identification would be negatively associated with evaluations of an undesirable ingroup member, we regressed the target evaluation scores on the standardized identification scores, the target variable, and the Identification  $\times$  Target interaction term.

The identification main effect was not significant ( $\beta = .11, t[63] = 1.29, p = .20$ ) but the target main effect ( $\beta = -.88, t[63] = -15.72, p < .001$ ) and the Identification  $\times$  Target interaction ( $\beta = -.28, t[63] = -3.40, p = .001$ ) were significant. The relationship between identification and the evaluation scores was negative and significant in the undesirable target condition ( $\beta = -.53, t[33] = -3.64, p = .001, R^2 = .29$ ) and positive but not significant in the desirable target condition ( $\beta = .23, t[30] = 1.28, p = .21, R^2 = .05$ ). This result showed that the more participants identified with the ingroup, the less favorable they were towards an undesirable ingroup member.

#### Mediation analyses

Finally, we predicted that identification would have an impact on stereotype change in response to the presentation of an undesirable ingroup member through evaluations of the undesirable ingroup member. To test this hypothesis, we conducted a series of regression analyses using data from the undesirable target condition,<sup>3</sup> following procedures outlined by Baron and Kenny (1986). The mediation analysis is represented graphically in Fig. 2. As well as the direct effects of identification on the target evaluations and stereotype ratings described above, the analyses also revealed that the target evaluations predicted the stereotype ratings ( $\beta = -.21, t[32] = -3.25, p = .003$ ), and the direct effect of identification on the stereotype ratings was reduced to a non-significant level ( $\beta = .14, t[31] = 1.53, p = .14$ ) when the target

evaluation scores ( $\beta = -.15, t[31] = -2.60, p = .007$ ) were also included as a predictor in the analysis. A Sobel (1982) test confirmed that the mediation was significant ( $z = 2.16, p = .003$ ).<sup>4</sup>

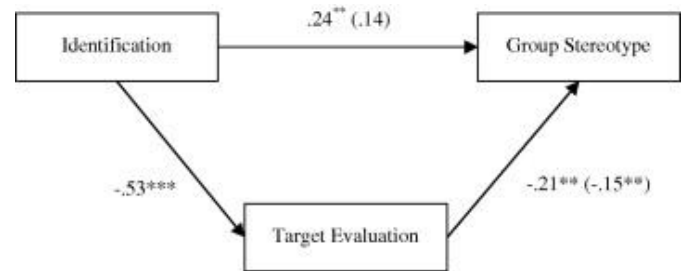


Fig. 2. Mediation analysis Study 1 (undesirable target condition). *Note.* Group stereotype ratings are the post-presentation ratings with the pre-presentation ratings controlled for. The numbers in parentheses are the  $\beta$ weights for the indirect effects. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

## Discussion

The results are consistent with predictions. The more participants identified with the ingroup, the more positive they rated the ingroup stereotype after reading about an undesirable ingroup member. This suggests that high identifiers may have excluded the undesirable target from their representation of the ingroup. This interpretation is supported further by the target evaluation ratings, which showed that identification was negatively associated with evaluations of the undesirable ingroup member. Furthermore, mediation analyses confirmed that identification had an impact on the stereotype ratings following presentation of an undesirable ingroup member: the stronger the identification, the more negative the evaluation, and the greater the shift towards a more positive ingroup stereotype.

## Study 2

The results from Study 1 suggest that devaluation might serve to exclude undesirable members from the representation of the ingroup. Study 2 aimed to replicate and extend these findings by also examining how the presentation of an undesirable outgroup member might impact on the perception of the outgroup stereotype. If devaluation indeed excludes undesirable members from the group, then we would expect high identifiers to be more motivated to exclude an undesirable target from the representation they form when judging the ingroup than from their representation of the outgroup. This difference, we reasoned, should be reflected in ingroup and outgroup stereotype ratings following presentation of an undesirable ingroup or outgroup member, respectively. Study 2 was designed to test this prediction.

One potential limitation to Study 1 was that measuring group ratings both before and after the presentation of a target group member may have created demand characteristics. To reduce this possibility in Study 2, rather than all participants completing the stereotype items twice, half completed the items before the presentation of a target group member and half after the presentation. Additionally, in Study 1 participants indicated their level of identification with the group 'psychology students' before evaluating a 'professional psychologist'. This raises the possibility that participants may not have shared an identity with the target group member, which would undermine our suggestion that devaluation of the undesirable target reflected an identity maintenance strategy. To reduce this possibility in Study 2, the participants and targets were more clearly members of a common category: university students.

## Method

### Participants

Participants were 180 students from the University of Kent. One hundred and forty-nine were female, 29 were male, and two did not specify their gender. Participants' ages ranged from 17 to 57 ( $M = 20.87$ ,  $SD = 5.99$ ).

### Materials

#### Cover story and instructions

Participants were informed that the purpose of the study was to assess how students from different universities perceive themselves and each other. They were told that students from the Universities of Kent (UKC) and Essex were participating. Pilot testing indicated that the universities are perceived by UKC students as similar in size and status and that the students are perceived in equally positive stereotypical terms. Specifically, students from both universities are perceived as equally hardworking, ambitious, dedicated, committed, professional, and responsible. Participants were informed that they would be asked to form an image of a target student in order to be able to predict and interpret his future behavior. Responses to all items were made on rating-scales ranging from 1 (*not at all*) to 9 (*very much*).

#### Group identification measure

Participants completed the same four *group identification* items from Study 1 to assess their level of identification with their university ( $\alpha = .82$ ).

#### Group stereotype measure

Six items measured the perception of UKC or Essex students on a series of positive and negative stereotypical dimensions: 'hardworking,' 'lack ambition,' 'dedicated,' 'lack commitment,' 'professional,' 'irresponsible.' The negative items were reverse scored and combined with the positive items to form a single *group stereotype* score ( $\alpha = .92$ ). A higher score indicates a more positive stereotype.

#### Target descriptions

Participants were presented with one of four descriptions of a target student. The student was either from the University of Kent (*ingroup condition*) or the University of Essex (*outgroup condition*). All four descriptions began with the same demographic information and continued to describe the target's attitude toward and commitment to his work, his ambitions, and so on. The target student either confirmed (*desirable target condition*) or disconfirmed (*undesirable target condition*) the positive stereotype of students. Examples of sentences used to describe the desirable student included the following: "He regularly attends lectures, never missed exams, and friends often rely on him for information," and "[He] has the intellectual ability to successfully complete the degree course and also the required dedication and motivation." When asked to comment on his outlook, the desirable student responded as follows: "I came here primarily to study, but also to have fun and meet new people. It is important to get the balance between work and play right."

In contrast, examples used to describe the undesirable student included the following: "He fails to attend lectures regularly, often missed exams, and relies on classmates for information," and "[He] has the intellectual ability to successfully complete the degree course but he lacks the required dedication and motivation." When asked to comment on his outlook, the undesirable student responded as follows:

"I came here to have fun and meet new people, not to spend my time studying. There are better things for young people to do with their time than read books."

#### Target evaluation measure

Three items measured participants' evaluation of the targets: "I like this student's outlook," "This person is a good student," "I would like to work with this student." These items were combined to form a single *target evaluation* score ( $\alpha = .87$ ).

#### Procedure

The study was conducted in a lecture room. Participants were randomly assigned to conditions. All participants first completed the identification items. Half of them completed the stereotype items before reading the target description. Those participants then evaluated the target. The other half read the target description, completed the stereotype items, and evaluated the target. Debriefing took place in the following lecture.

## Results

### Group stereotype ratings

We predicted that identification with the ingroup would correlate positively with ingroup stereotype ratings following presentation of an undesirable ingroup member. We tested this prediction using hierarchical regression analyses. In four steps we entered (1) the predictors *group* (effect-coded -1 = ingroup, 1 = outgroup), *target* (effect-coded -1 = desirable, 1 = undesirable), *time of judgment* (effect-coded -1 = pre-manipulation, 1 = post-manipulation) and the standardized *group identification* scores, (2) the pairwise products of these predictors, (3) the three-way interaction terms, and (4) the product of all predictors to test the four-way interaction.

The four-way interaction was significant ( $\beta = -.48$ ,  $t[164] = -2.08$ ,  $p = .039$ , all other  $ts < 1.36$ ,  $ps > .17$ ). To test our specific hypotheses we compared the slopes of the regression lines for identification across all combinations of group and time of judgment in both target conditions (see Fig. 3). The slopes were not significant in the desirable target condition (all  $ts < 1.16$ ,  $ps > .26$ ). In the undesirable target condition, identification was positively associated with ingroup stereotype ratings after ( $\beta = .40$ ,  $t[21] = 2.01$ ,  $p = .06$ ,  $R^2 = .16$ ), but not before ( $\beta = .02$ ,  $t[18] = .07$ ,  $p = .95$ ,  $R^2 < .001$ ), the presentation of an undesirable ingroup member. In contrast, identification was negatively associated with outgroup stereotype ratings after ( $\beta = -.48$ ,  $t[21] = -1.98$ ,  $p = .061$ ,  $R^2 = .16$ ), but not before ( $\beta = .09$ ,  $t[22] = .41$ ,  $p = .69$ ,  $R^2 = .01$ ), the presentation of an undesirable outgroup member. These results indicated that the more participants identified with the ingroup, the more positive they rated the ingroup stereotype following the presentation of an undesirable ingroup member, and the more negative they rated the outgroup stereotype after reading about an undesirable outgroup member.

The target evaluation ratings were also analyzed using hierarchical regression analysis. In three steps we entered (1) the independent variables, (2) their cross products, and (3) the three-way interaction of Identification  $\times$  Target  $\times$  Group. The analysis revealed a significant main effect of target ( $\beta = -.84$ ,  $t[172] = -23.33$ ,  $p < .001$ ), a significant Group  $\times$  Identification interaction ( $\beta = .08$ ,  $t[172] = 2.31$ ,  $p = .02$ ), and a significant Group  $\times$  Target interaction ( $\beta = .17$ ,  $t[172] = 4.83$ ,  $p < .001$ ).

These effects were qualified by a significant Identification  $\times$  Group  $\times$  Target interaction ( $\beta = .16, t[172] = 4.54, p < .001$ ; all other  $t$ s  $< 1, p$ s  $> .45$ ).

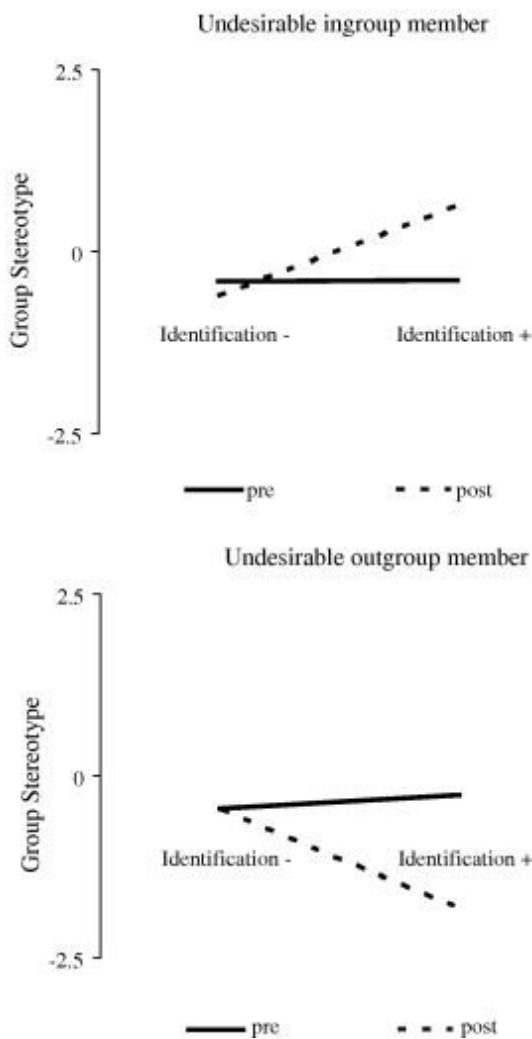


Fig. 3. Graphical representation of interaction between identification, group, target, and time of judgment on group stereotype ratings.

#### Target evaluation ratings

To establish whether the familiar black sheep effect emerged (Marques & Yzerbyt, 1988), we decomposed the significant Group  $\times$  Target interaction by comparing the slopes of the regression lines for group in both target conditions. The slope was negative in the desirable target condition ( $\beta = -.31, t[88] = -3.10, p = .003, R^2 = .10$ ) and positive in the undesirable target condition ( $\beta = .34, t[88] = 3.33, p = .001, R^2 = .11$ ). This indicated that participants favored a desirable ingroup member over a desirable outgroup member, but favored an undesirable outgroup member over an undesirable ingroup member. Thus, the black sheep effect was established.

To test our prediction that identification with the ingroup would predict evaluations of desirable and undesirable ingroup and outgroup members, we decomposed the significant Identification  $\times$  Group  $\times$  Target interaction by comparing the slopes of the regression lines for identification across all combinations of group and target. Identification was positively associated with evaluations of the desirable ingroup member ( $\beta = .31, t[46] = 2.23, p = .03, R^2 = .10$ ), whereas the relationship between identification and evaluations of the desirable outgroup member was not significant ( $\beta = -.08, t[40] = -.51, p = .62, R^2 = .01$ ). In contrast,

identification was negatively associated with evaluations of the undesirable ingroup member ( $\beta = -.56, t[41] = -4.28, p < .001, R^2 = .31$ ), whereas the relationship between identification and evaluations of the undesirable outgroup member was positive but not significant ( $\beta = .25, t[45] = 1.70, p = .10, R^2 = .06$ ). This confirmed that the more participants identified with the ingroup, the more favorable they were towards a desirable ingroup member and the less favorable they were towards an undesirable ingroup member.

#### Mediation analyses

As in Study 1, we predicted that the relationship between ingroup identification and ratings of the ingroup stereotype following the presentation of an undesirable ingroup member would be mediated by evaluations of the undesirable ingroup member. To test this prediction we conducted a series of regression analyses using data from the ingroup undesirable target condition (Baron & Kenny, 1986).<sup>5</sup> The mediation analysis is represented graphically in Fig. 4. As well as the above-described effects of identification on the target evaluations and group stereotype ratings, the regression analyses also confirmed that the target evaluations predicted the stereotype ratings ( $\beta = -.53, t[21] = -3.73, p = .001$ ), and the direct effect of identification on the stereotype ratings became non-significant ( $\beta = .09, t[20] = .45, p = .65$ ) when the target evaluation scores ( $\beta = -.48, t[20] = -2.86, p = .01$ ) were included in the analysis. A Sobel test confirmed that the mediation was significant ( $z = 2.03, p = .04$ ).<sup>6</sup>

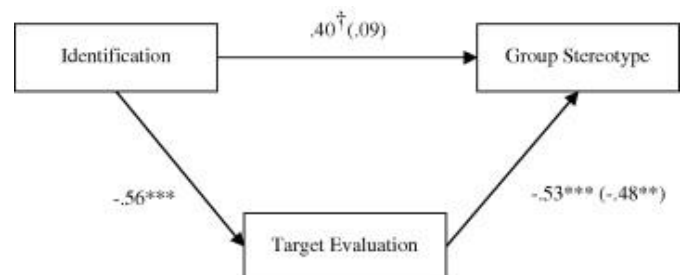


Fig. 4. Mediation analysis Study 2 (undesirable ingroup member condition). Note. The numbers in parentheses are the  $\beta$  weights for the indirect effects. \*\*\* $p < .001$ , \*\* $p < .01$ , † $p = .06$ .

#### Discussion

The results are again consistent with predictions. Participants favored a desirable ingroup member over a desirable outgroup member, but favored an undesirable outgroup member over an undesirable ingroup member, consistent with the black sheep effect (Marques & Yzerbyt, 1988). Additionally, identification with the ingroup was positively associated with ingroup stereotype ratings following the presentation of an undesirable ingroup member but was negatively associated with outgroup ratings following presentation of an undesirable outgroup member. As in Study 1, the relationship between ingroup identification and ingroup stereotype ratings was mediated by evaluations of the undesirable ingroup member: the stronger the identification, the more negative the evaluation, and the greater the shift towards a more positive perception of the ingroup stereotype.

#### General discussion

Our starting point was the assumption, following traditional explanations of the black sheep effect (e.g., Marques & Paez, 1994), that high identifiers would be motivated to exclude an undesirable member from their representation of the ingroup. We reasoned that excluding an undesirable member from the ingroup would have a contrast effect on subsequent perceptions

of the ingroup stereotype. Consistent with this, in both studies we found that ingroup identification had an impact on the ingroup stereotype ratings following the presentation of an undesirable ingroup member through the target evaluations ratings.

In Study 2, participants were more negative towards an undesirable ingroup member than an identical outgroup member, and only when the undesirable target was an ingroup member did the target evaluation scores mediate the impact of identification on the stereotype ratings. Thus, although the undesirable ingroup and outgroup members were devalued relative to their desirable counterparts, it seems that devaluation was related to psychological exclusion only when the undesirable target was an ingroup member. When the undesirable target was an outgroup member, the stronger the ingroup identification, the greater was the shift towards a more negative outgroup stereotype. This suggests that high identifiers were more motivated to include an undesirable target in the representation they formed when judging the outgroup. It may be that including an undesirable member in the representation of the outgroup allows high identifiers to establish or maintain a positive distinction between the ingroup and the outgroup (see [Abrams, Marques, Randsley de Moura, Hutchison, & Bown, 2004](#)).

The current research extends our understanding of the social cognitive processes that might underlie the familiar black sheep effect ([Marques & Yzerbyt, 1988](#)). The results suggest that high identifiers were more extreme in their devaluation of an undesirable ingroup member because they were more motivated to exclude that member from their representation of the ingroup. If the negative evaluations were not driven by a motivation to exclude the undesirable target from the ingroup, then there is no reason why the target evaluation ratings should have mediated the effect of identification on the ingroup stereotype ratings. Our results suggest that high identifiers were indeed motivated to exclude the undesirable target from the ingroup, and so the more they devalued the undesirable ingroup member, the more positive they perceived the ingroup stereotype.

Although the results are consistent with our predictions, it is important to address some potential limitations with the research. One concern is the possibility that the results may have been biased by demand characteristics. This may be a concern especially in Study 1, in which participants completed the same stereotype items both before and after reading about a target group member. However, there are various reasons to believe that demand characteristics did not influence the results. First, the presentation of a desirable target had no effect on the group ratings. A demand account would have to explain why the presentation of an undesirable target would produce a demand but not the presentation of a desirable target. Second, only high identifiers modified their perceptions of the group stereotype following the presentation of an undesirable group member. A demand account would have to explain why participants with different levels of identification would react differently to what would appear as a demand of the experiment. Third, similar results were obtained in Study 2, in which half the participants completed the stereotype items before the presentation of a target group member and half after the presentation. Finally, none of the participants expressed any accurate suspicions about the true aims of the research. For these reasons, we are confident that demand characteristics did not bias the results. Nevertheless, future research should aim to replicate the effects using different paradigms.

It should also be noted that in both studies, the level of ingroup identification did not predict the group ratings taken before the presentation of the target group members or following the presentation of a desirable ingroup or outgroup member. This may seem at odds with the assumption that high identifiers will use every opportunity to perceive the ingroup in a positive light. However, the results align with previous research showing that it is only when the ingroup is threatened that the level of identification with the ingroup becomes a reliable predictor of group-based judgments and evaluations (e.g., [Hutchison, Jetten, Christian, & Haycraft, 2006](#)).

Additionally, like previous research on the black sheep effect (see [Marques & Paez, 1994](#)), the ‘deviants’ in our research were intentionally portrayed as undesirable and unfavorable. However, several studies have suggested that perceptions of deviance can change across different social contexts, to the extent that an individual that is seen as being deviant in one context need not necessarily be deemed deviant in another context (e.g., [Matheson, Cole, & Majka, 2003](#)). An interesting avenue for future research is to assess how the same target person might impact on perceptions of their group in different comparative contexts.

In conclusion, the results from two studies are consistent with traditional interpretations of the black sheep effect, which have argued that in derogating undesirable ingroup members, high identifiers would attempt to protect the ingroup by “getting rid of the bad ones” ([Yzerbyt, Castano, Leyens, & Paladino, 2000, p. 267](#)). While this interpretation has received considerable circumstantial support from research comparing evaluations of desirable and undesirable ingroup and outgroup members, the current research is the first to establish empirically the relationship between group identification, deviant derogation, and identity maintenance.

## References

- Abrams, D., Marques, J. M., Randsley de Moura, G., Hutchison, P., & Bown, N. J. (2004). The maintenance of entitativity: A subjective group dynamics approach. In V. Y. Yzerbyt, C. M. Judd, & O. Corneille (Eds.), *The psychology of group perception: Contributions to the study of homogeneity, entitativity and essentialism* (pp. 361–380). Philadelphia, PA: Psychology Press.
- Aitken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Bless, H., & Schwarz, N. (1998). Context effects in political judgement: Assimilation and contrast as a function of categorization process. *European Journal of Social Psychology*, 28, 287–291.
- Brewer, M. B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86, 307–324.
- Castano, E., Paladino, M.-P., Coull, A., & Yzerbyt, V. Y. (2002). Protecting the ingroup stereotype: Ingroup identification and the management of deviant ingroup members. *British Journal of Social Psychology*, 41, 365–385.
- Coull, A., Yzerbyt, V. Y., Castano, E., Paladino, M.-P., & Leemans, V. (2001). Protecting the ingroup: Motivated allocation of cognitive resources in the presence of threatening

ingroup members. *Group Processes and Intergroup Relations*, 4, 327–339.

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, 31, 410–436.

Eidelman, S., Silvia, P. J., & Beirnet, M. (2006). Responding to deviance: Target exclusion and differential devaluation. *Personality and Social Psychology Bulletin*, 32, 1153–1164.

Hutchison, P., & Abrams, D. (2003). Ingroup identification moderates stereotype change in reaction to ingroup deviance. *European Journal of Social Psychology*, 33, 497–506.

Hutchison, P., Jetten, J., Christian, J., & Haycraft, E. (2006). Protecting threatened identity: Sticking with the group by emphasizing ingroup heterogeneity. *Personality and Social Psychology Bulletin*, 32, 1–13.

Kunda, Z., & Oleson, K. C. (1997). When exceptions prove the rule: How extremity of deviance determines the impact of deviant examples on stereotypes. *Journal of Personality and Social Psychology*, 72, 965–979.

Marques, J. M., & Paez, D. (1994). The 'black sheep effect': Social categorization, rejection of ingroup deviates, and perception of group variability. In W. Stroebe & M. Hewstone (Eds.). *European review of social psychology* (Vol. 5, pp. 37–68). Chichester, UK: Wiley.

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

Matheson, K., Cole, B., & Majka, K. (2003). Dissidence from within: Examining intragroup reactions to attitudinal opposition. *Journal of Experimental Social Psychology*, 39, 161–169.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, 22, 103–122.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. *Sociological methodology 1982* (pp. 290–312). San Francisco, CA: Jossey-Bass.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.

Yzerbyt, V. Y., Castano, E., Leyens, J.-P., & Paladino, M.-P. (2000). The primacy of the ingroup: The interplay of entitativity and identification. In W. Stroebe & M. Hewstone (Eds.). *European review of social psychology* (Vol. 11, pp. 257–296). Chichester, UK: Wiley.96

## Notes

1 An accurate suspicion was defined a priori as one suggesting that the aim of the research was to assess how exposure to a target group member might influence group ratings (or vice-versa). The majority of participants reported no suspicions.

2 When analyzing pre-test/post-test data in this way it is important to test the assumption of homogeneity of regression. This is the assumption that the slopes of the regression of the

dependent variable on the covariate (pre-presentation group stereotype ratings) are the same for all combinations of the independent variables (identification and target). If the covariate does not interact with the independent variables, then the assumption has not been violated. None of the interactions involving pre-presentation stereotype ratings was significant. Additionally, there were no apparent non-normalities in the residuals and no multivariate outliers.

3 We did not conduct a mediation analysis using data from the desirable target condition because the conditions for mediation to be present were not satisfied in that condition.

4 The mediation analysis suggests that differences in the group stereotype ratings associated with the level of identification were a result of evaluations of the undesirable target. However, it is theoretically possible that differences in the group ratings associated with identification impacted on the target evaluations. That is, the presentation of an undesirable ingroup member may have had an impact on the representation of the group, which in turn may have led to harsher treatment of the target. To explore this possibility we tested an alternative model which reversed the roles of the target evaluations and group stereotype ratings. In this model identification was the independent variable, target evaluation was the dependent variable, and the group stereotype score was the mediator. The data did not support this model. After controlling for the stereotype scores the effect of identification on the target evaluation scores remained significant ( $\beta = -.44$ ,  $t[31] = -2.23$ ,  $p = .03$ ) and a Sobel test indicated that mediation had not occurred ( $z = 1.67$ ,  $p = .10$ ).

5 As in Study 1, we conducted a mediation analysis on data from the undesirable condition only because the conditions for mediation to be present were not satisfied in the other conditions.

6 As in Study 1, we tested an alternative model in which the roles of the target evaluations and group stereotype ratings were reversed. The data did not support this model. After controlling for the stereotype scores, the effect of identification on the target evaluation scores remained (marginally) significant ( $\beta = -.33$ ,  $t[20] = -1.90$ ,  $p = .07$ ) and a Sobel test indicated that mediation had not occurred ( $z = 1.64$ ,  $p = .10$ ).