Not as black as one of 'them':

Positive ingroup valence reverses the Black Sheep Effect

Johann Jacoby
Knowledge Media Research Center, Tübingen, Federal Republic of Germany

Sabine Otten
Rijksuniversiteit Groningen, The Netherlands

Kai Sassenberg
Knowledge Media Research Center, Tübingen, Federal Republic of Germany

8206 words (all incl.)
Abstract

Judgments about individual group members are often colored by judgments about the entire group. Through such an assimilation effect, an individual norm violator from a positively evaluated group, for example an ingroup, can be expected to be evaluated less negatively than an outgroup norm violator. This hypothesis however contrasts with evidence that ingroup norm violators are often derogated compared to outgroup norm violators, a phenomenon known as the Black Sheep Effect (BSE). The discussion of differences between the explanations of these opposite predictions (assimilation vs. BSE) lead us to the prediction that assimilation (and not a BSE) occurs if the ingroup adheres more to the societal norm than the outgroup and thus conveys positive valence to which an ingroup deviant evaluation is assimilated. This predicted pattern was indeed found in Study 1 and replicated in Study 2. Additionally, Study 2 provides evidence that this effect is, consistent with the present theorizing, mediated by group valence. The current research suggests that a BSE can be eliminated or even reversed by assimilation effects – a suggestion that is in line with other differentiations in the literature regarding the circumstances under which a BSE, no difference, or even ingroup favoritism occurs.

(195 words)

Keywords: intergroup bias, Black sheep effect, norm violation, valence assimilation
It is well known from person perception and stereotyping research that perceivers tend to use information in judging individual targets that actually does not strictly pertain to the target and her behavior (Fiske & Neuberg, 1990; Wyer & Srull, 1989). For example, perceptions of a target's ambiguous behaviors are colored by independently activated concepts (e.g., Devine, 1989) or a target's race (Sommers & Ellsworth, 2000; Willis Esqueda, 1997). Such assimilation effects are especially likely if no other information about a target is available and the information about an individual is ambiguous (cf. the Donald paradigm, see Higgins, Rholes, & Jones, 1977).

Based on this research, one could expect that a target showing undesirable behavior may be judged differently depending on her group membership. If that group's image is generally positive, that target's evaluation should be relatively less negative than if her group is viewed negatively.

However, research from a social identity perspective has repeatedly found a pattern that seems to be at odds with this prediction: the *Black Sheep Effect* (BSE, Marques & Paéz, 1994). Given that ingroups are in general viewed as more positive than outgroups (Hewstone, Rubin, & Willis, 2002; Tajfel, Billig, Bundy, & Flament, 1971; Tajfel & Turner, 1986), it is somewhat surprising at first glance that this research found consistent evidence that ingroup norm violators are judged more negatively than outgroup norm violators (Marques & Páez, 1994).

In the presently reported research we propose and test an explanation for this contradiction between the assimilation hypothesis and the BSE. We propose that under certain circumstances, the process of assimilation takes precedence over social identity concerns and that consequently, the BSE is reversed: an ingroup norm deviant is judged less negatively than an outgroup deviant.
Assimilation effects from categories

Person perception models posit that judgments about an individual can incorporate information from potentially diverse sources (Moskowitz, 2005; Kruglanski & Sleeth-Keppler, 2007; Wyer & Srull, 1989). Some of that information can be objectively unrelated to the individual target proper, but stem from either incidentally activated concepts (such as in Higgins et al., 1977) or concepts that are associated with a target, for example a social category that the target belongs to (e.g., stereotyping, see Devine, 1989; Lepore & Brown, 1997; Wittenbrink, Judd, & Park, 1997). Such unrelated information can facilitate and possibly bias the identification (e.g., the shooter bias, Correll, Park, Judd, & Wittenbrink, 2002) and categorization (Darley & Gross, 1983; Hastorf & Cantril, 1954; Blair, Judd, & Chapleau, 2004) of a target and her behavior. These conceptual assimilation effects based on social categories require pre-existing associations, or developed stereotypes, which allow for the often automatic activation of particular content via the social category information about a target.

In addition to concepts, general valence can also transfer onto an individual target even if it is not pertinent to that target (Wittenbrink, Judd, & Park, 1997), a fact that is central to the mechanisms underlying particular indirect measures of attitudes (see the Affect Misattribution Paradigm, Payne, Cheng, Govorun, & Stewart, 2005). In this process, the evaluative judgment about an individual target is assimilated to an evaluative judgment about a different object, such as a group which the target belongs to. In line with research on evaluative conditioning (de Houwer, Thomas, & Baeyens, 2001; Walther, Nagengast, & Trasselli, 2005; Olson & Fazio, 2001) we argue here that such general valence associations can develop quickly and may even override concept-based associations that are the basis for the emergence of assimilation effects in
identification and categorization. If this is so, then the mere evaluation of an individual
target should be more easily affected by manipulated valence-relevant social category
information than by conceptual identification and categorization of a particular behavior
a target has shown.

From these considerations it is to be expected that an individual target profits
from belonging to a category that is associated with positive valence: a judgment about
her positive (or negative) behavior will be more positive (or less negative) than if
category membership is not known. If the category, on the other hand, is less associated
with positive valence, her actions should be judged less positively and more negatively,
respectively.

The Black Sheep Effect

A large number of studies have found the BSE since its first description (Marques,
Yzerbyt, & Leyens, 1988; Marques & Yzerbyt, 1988, for reviews, see Marques & Páez,
1994, or Marques, Abrams, Páez, & Hogg, 2001). This effect consists in a relatively
less positive evaluation of an ingroup as compared to an outgroup norm violator. It is
often accompanied by simultaneous ingroup favoritism in the evaluation of norm
abiding ingroup vs. outgroup targets (e.g., Abrams, Rutland, Cameron, 2003;
Branscombe, Wann, Noel, & Coleman, 1993).

In earlier theoretical treatments, derogating an ingroup deviant was interpreted as
an act of symbolic purging him or her from the ingroup such that the remaining
membership could still provide for an overall positive view of the ingroup (Marques &
Páez, 1994). Later theoretical developments refined and extended this explanation for
the BSE: The Subjective Group Dynamics Model (SGD, Abrams, Marques, Bown, &
Henson, 2000; Abrams, Marques, Randsley de Moura, Hutchinson, & Bown, 2005;
Abrams, Randsley de Moura, Hutchinson, & Viki, 2005) distinguishes between anti-norm deviants, that is "group members who reject their group's norms and/or who favor the normative attitudes or behavior of a contrasting group" (Abrams et al., 2000, p. 906) and pro-norm deviants who over-confirm group norms and thereby increase the pre-existing contrast between the ingroup and an outgroup. An ingroup and an outgroup deviant may thus show identical behavior, but the former is an anti-norm deviant relative to the ingroup's norm (i.e., undermines the ingroup norm and favors the outgroup's norm) while the latter is anti-norm deviant relative to the outgroup's norm and thereby "boosts relative validity" (Abrams et al., 2000) of the ingroup's norm. Therefore, the ingroup deviant is downgraded relative to the outgroup deviant even though both show the same behavior. Thus, within research on the BSE, derogating ingroup deviants is viewed and explained as a social identity protection strategy (Abrams et al., 2000).

The quote above that defines anti-norm deviance, shows that the SGD presupposes a contrast of the ingroup and an outgroup in norms. Also, prominent instances of the BSE in the literature have occurred for violations of norms for which a difference between the entire ingroup vs. outgroup exists in the first place, but where this pre-existing norm difference was threatened to be eliminated or reduced by the deviants’ actions (Marques et al., 1988, Experiment 2). Other studies have revealed that the BSE is eliminated if the difference between ingroup and outgroup is secure (Marques et al., 2001, Experiment 3) or not salient (Marques et al., 2001, Experiment 2). These results indicate that the BSE is a reaction to a threat to an already established difference in norms between ingroup and outgroup that, to perceivers, could however relatively easily be eliminated by individual members’ norm deviance. In this instance
of threat, the positive social identity drawn from membership in the ingroup is at risk and therefore, presumably, participants reaffirm the ingroup's superiority by derogating ingroup deviants.

On the basis of the summarized research on the BSE and the theorizing accompanying it, one would therefore generally expect an ingroup deviant to be derogated compared to an outgroup deviant especially if already established ingroup superiority is at risk to be blurred by an ingroup deviant. This pattern directly contradicts the prediction of an assimilation effect.

When assimilation overrides social identity concerns

We propose that the derogation of a deviating member of a positively viewed ingroup will not emerge if a norm violation is an actual behavior that does not only defy group norms, but superordinate norms. The negative affect elicited by this behavior does not fundamentally threaten ingroup identity, but superordinate values and rules. We predict that then a difference in norms between the ingroup and an outgroup as a whole will rather color individual deviant evaluations: individual judgments will be assimilated to the respective group judgments. This should result in an ingroup deviant being evaluated less negatively than an outgroup deviant showing the same behavior.

We propose that social identity concerns, which figure prominently as a driving force in the BSE, are weaker if the negative affect elicited by the behavior in question is in general clearly negative. The overwhelming majority of BSE have been found using deviant behavior in relation to certain attitudes or personal preferences (e.g., recommended percentage of increase or decrease in acceptance of Asylum seekers in Abrams et al., 2000; attitude towards student initiation practices, Marques et al., 2001; attitude towards preferential treatment of overseas students in campus housing policy,
Abrams et al., 2002). Such deviating attitudes or preferences do not carry strong negative valence per se in a liberal, pluralistic society. They are associated with negative valence only insofar as they strongly diverge from opinions and values seen as central to and cherished by one’s own group (see also Marques, 1990, Study 5).

Norm violating behaviors that are already viewed negatively independently of group membership in contrast, carry a relatively strong and inherent negative valence. For example, fraud is viewed negatively in society by itself already, independently of a defrauder's group membership. However, information about the prevalence of fraud among a group as a whole will change the evaluation of that group. This group evaluation should, in turn, have an assimilative effect on individual defrauders. Since affect plays a prominent role in judgments of transgressing behavior and often trumps deliberative, and rational elements (see Haidt & Kesebir, 2010), we expect it to override the concerns for group norm legitimacy germane to group-specific attitudes and preferences pertinent to the SGD. We predict that this affect thus hinders the process that under different circumstances would lead to a BSE and will give way to an assimilation effect.

Overview over the current studies

In order to test this prediction, we presented participants with an ingroup or an outgroup target who had shown an actual behavior violating a general norm (i.e., a college student who had cheated in an exam, an instance of fraud that is illegal in all universities). In addition, the ingroup and outgroup were either presented as differing as a whole in adhering to that norm (different adherence condition, the ingroup is portrayed as adhering to the norm more than the outgroup) or as not differing (equal adherence condition). An assimilation effect should be apparent in form of a relatively
more favorable evaluation of the ingroup target (compared to the outgroup target) in the different adherence condition, while no difference is expected under equal adherence. In contrast, based on the BSE literature one would expect derogation of the ingroup target in both conditions. In the different adherence condition, pre-existing ingroup superiority is threatened, and in the equal adherence condition ingroup superiority is non-existent and therefore participants should be motivated to create it through symbolic exclusion of the ingroup deviant by derogation.

While Study 1 tests these competing hypotheses and establishes the assimilation pattern, in Study 2 we replicate this result even under conditions increasing the threat to ingroup superiority and additionally test the hypothesis that group evaluations mediate individual person evaluations that is implied by the assimilation account.

Study 1

Method

Participants and design

Participants were 80 students of Friedrich Schiller University Jena in the state of Thuringia of the Federal Republic of Germany (mean age: 22 years, range: 18-30 years, 54% female). They volunteered to fill out a questionnaire on campus, received a chocolate bar for compensation and were later debriefed by email. Assignment to the cells of a 2 (Group adherence: equal vs. different) × 2 (Target’s group membership: ingroup vs. outgroup) between participants study design was random.

Procedure

Participants completed a questionnaire package containing all materials. They first read a fabricated article from an alleged online portal for university students about cheating in university exams. This article reported a survey among university students
in Germany about their opinions on cheating and their own past cheating behavior. In all conditions, the article stated that on average, 12% of students in Germany had admitted having cheated in an exam at least once. *Group adherence* was manipulated through an illustrative example of results from the survey either being overall very homogeneous across the German states (*equal adherence condition*) or varying considerably among the states (*different adherence condition*). In the former, equal adherence condition, the report stated that 12.1% in the neighboring state of Saxonia and 12% of students in the state of Thuringia had admitted having cheated in an exam. Also, it was reported that 71.8% of the students in Saxonia and 73.1% of students in Thuringia had indicated that they categorically rejected the option of cheating in the service of obtaining good grades. In the different adherence condition on the other hand, the percentages were 22.6% (outgroup) and 3.4% (ingroup) for admitting having cheated and 58.7% (outgroup) and 83.2% (ingroup) for rejecting the option of cheating. The information on students categorically rejecting cheating was added to curb the possibility that participants assumed that the respondents who had not admitted to cheating had in fact cheated, but lied in the survey.

In the last paragraph of the article, a student was quoted who admitted that he had cheated using a mobile phone. To manipulate *Target’s group membership*, this student was parenthetically introduced as a student of either Technical University Dresden (in Saxonia, *outgroup target* condition) or a student of Friedrich Schiller University Jena (Thuringia, *ingroup target* condition).

After reading the article, participants completed the following dependent measures.

*Dependent measures*
Participants were asked to indicate their impression of the quoted student regarding 11 adjective pairs on seven point bipolar rating scales with higher values indicating more negative evaluation (e.g., pleasant – unpleasant, good – bad, reliable – unreliable). As these items formed a reliable scale (α = .87) they were averaged to an index of person evaluation.

Participants also rated how well each of five adjectives fit the student's behavior (false, scandalous, unacceptable, outrageous, mean) on seven point scales anchored by 1='does not fit at all' and 7='fits very well'. These items were averaged for an index of behavior evaluation (α = .85). This measurement was taken to protect against the possibility that participants' reactions to the target person could be explained via different interpretations of the same behavior (e.g., as witty and ingenious in the ingroup condition vs. despicable and wretched in the outgroup condition) rather than evaluation of the target person. But we predict ingroup favoritism in the evaluation of norm violators, and a different interpretation of the behavior itself could jeopardize unambiguous assessment of these target student evaluations. Thus, while our hypothesis does predict differences in person evaluations, we did not expect differences in behavior evaluation as a function of Group membership or Group adherence.

Results

Our prediction was that the ingroup target would be evaluated less negatively than the outgroup target in the different adherence condition, but no difference would emerge in the equal adherence condition. An alternative prediction, based on the SGD, is that the ingroup target would be evaluated more negatively than the outgroup target in any or both conditions.

To test these hypotheses, we conducted 2 (Group adherence: equal vs. different) ×
2 (Target’s group membership: ingroup vs. outgroup) ANOVAs on person evaluation and behavior evaluation, respectively. Person evaluation and behavior evaluation were moderately correlated, \( r(80) = .34, p = .002 \). However, result patterns for these two dependent variables were different regarding the pertinent Group adherence × Target’s group membership interaction, therefore separate analyses are reported. Means and standard deviations are shown in Table 1.

For person evaluation no main effect of Group adherence was found, \( F(1,76) = 2.42, p = .124, \eta_p^2 = .03 \), nor an effect of Target’s group membership by itself, \( F < 1 \), but a Group adherence × Target’s group membership interaction, \( F(1,76) = 7.37, p = .008, \eta_p^2 = .09 \). In the equal adherence condition, the difference between the targets was not significant (\( M_{ig} = 4.90, SD = 0.86; M_{og} = 4.50, SD = 0.62 \)), \( F(1,76) = 2.67, p = .107, \eta_p^2 = .03 \). In the different adherence condition, however, the ingroup target was rated distinctly less negatively (\( M_{ig} = 4.16, SD = 0.90 \)) than the outgroup target (\( M_{og} = 4.70, SD = 0.71 \)), \( F(1,76) = 4.87, p = .03, \eta_p^2 = .06 \). Thus, in line with our predictions, there was a clear assimilation effect in the different adherence condition and not a BSE. No significant effect emerged in the equal adherence conditions.

As expected, no effects emerged for behavior evaluation scores, all \( Fs < 1 \). The evaluation of the behavior itself was thus not affected by the manipulations. Hence, the effect found for person evaluation cannot be explained by different interpretations of the behavior itself, and is specific to the target person.

Discussion

As expected from an assimilation perspective, no significant difference between person evaluations was found if the two groups were portrayed as equally adhering to the norm in question, even though descriptively, the results hinted at a tendency for a
BSE. However, when adherence was higher in the ingroup – and a pre-established ingroup superiority was truly threatened by the ingroup deviant – this difference lead to an assimilation of individual judgments to group valence instead of a BSE. In this case, the ingroup target was favored. The fact that there were no differences in evaluations of the behavior itself shows that the effect is indeed specific to the target person and not due to a different interpretation of the objectively identical target behaviors.

Study 2

Above we hypothesized that an ingroup norm violator would profit from a positive image of the entire ingroup. This implies that by manipulating different cheating rates in the groups, the valence of the ingroup presented as cheating less should be perceived as more positive than that of the outgroup as a whole. Different group evaluations should therefore mediate the ingroup favoritism effect in person evaluations in the different adherence condition. Following the SGD, however, the overall group difference in norm adherence should make the concern with ingroup positivity (that could be lost) more pressing and rather lead to a BSE. Even though such an ingroup derogation effect was not visible in Study 1, it may have been present but smaller than the assimilation effect and thus been overruled by the assimilation effect. If this is so, statistically controlling for the effect of group valence in the course of mediation analysis should not only eliminate the assimilation effect in target evaluations (as predicted by the assimilation account), but even reveal a residual BSE.

In the next study, we therefore included measures of global ingroup and outgroup evaluations as mediator variable candidates. If the above ideas are viable, different group valence as a consequence of the manipulation of different group adherence should be discernible on these measures, and these differences should in turn be related
to individual target judgments. Additionally, statistically controlling for group evaluations as a covariate a) should eliminate the ingroup favoritism effect found in the different adherence condition in Study 1 and b) might even reveal a BSE (smaller than – and thus masked by – the ingroup favoritism effect) in person evaluations.

Moreover, we increased the value that the positive ingroup distinctiveness could represent for participants, thereby aiming at increasing the relevance of the group difference and thus the potential threat posed by an ingroup deviant to ingroup positivity. Following the SGD, tendencies towards a BSE should become stronger under these circumstances (see Marques, 1990, Study 5; Marques et al., 1988, Study 2). Thus, if a BSE was indeed operating in Study 1, but was overlaid by the assimilation effect, it has a fairer chance in Study 2 to emerge because of this increased potential for identity threat, especially when the assimilation effect is controlled for in a mediation analysis.

Method

Participants and Design

Ninety-four students of Fachhochschule Jena (University of Applied Sciences) participated in this study. As in Study 1, participants volunteered to participate on campus for a bar of chocolate as compensation and were debriefed by email after data collection was completed. Group adherence (equal vs. different) and Target’s group membership (ingroup vs. outgroup) were again manipulated orthogonally between participants. Participants were on average 23 years old (range: 18-42 years), 38% were female.

Procedure

Participants completed the same questionnaire as in Study 1 with two modifications: in the different adherence condition, in addition to the differences in
norm adherence, the article mentioned that the cheating rate in Thuringia (ingroup) was also the lowest among all states in Germany. This should enhance the value of the difference between ingroup and outgroup and increase the threat posed by an ingroup deviant. Thus possible processes leading to a BSE should be strengthened and increase the detectability of such an effect, if it were indeed operating simultaneously with the assimilation effect that we predicted and found in Study 1.

The same variables as in Study 1 were measured: person evaluation ($\alpha = .80$) and behavior evaluation ($\alpha = .72$). Additionally, after evaluating the target and his behavior, participants were asked to give a global evaluation of the entire ingroup and outgroup on the same adjectives and scales that were used for the person evaluation. These evaluations formed reliable scales ($\alpha_s = .87$ and $.90$ for ingroup and outgroup, respectively), and were therefore averaged for indices of ingroup valence and outgroup valence, respectively.

**Results**

As in Study 1, we tested whether an assimilation effect (i.e., ingroup favoritism) emerged in the different adherence condition while no difference was apparent in the equal adherence condition. Descriptive statistics are shown in Table 2. A 2 (Group adherence: equal vs. different) $\times$ 2 (Target’s group membership: ingroup vs. outgroup) between participants ANOVA revealed a marginal main effect of Target’s group membership, $F(1,90) = 3.01$, $p = .086$, $\eta_p^2 = .03$, which was qualified by a Group adherence $\times$ Target’s group membership interaction, $F(1,90) = 9.99$, $p = .002$, $\eta_p^2 = .10$. As predicted and consistent with Study 1 there was no difference between the ingroup and the outgroup targets in the equal adherence condition, $M_{ig} = 4.44$, $SD = 0.73$; $M_{og} = 4.20$, $SD = 0.87$; $F(1,90) = 1.02$, $p = .316$, $\eta_p^2 = .01$, but when ingroup and outgroup
were presented as differing in norm adherence the ingroup target \( (M_{ig} = 3.86, SD = 0.76) \) was evaluated less negatively than the outgroup target \( (M_{og} = 4.68, SD = 0.87) \), \( F(1,90) = 11.98, p = .001, \eta_p^2 = .12 \). No Group adherence main effect emerged on person evaluation, \( F < 1 \). Again, the manipulations had no discernible effects on behavior evaluation, all \( Fs(1,90) < 1.02 \), all \( ps > .316 \), all \( \eta_p^2 s < .02 \).

Thus, the results of Study 1 were replicated: we found an ingroup favoritism effect on person evaluation in the different adherence condition, and not a BSE. No difference in person evaluation emerged when group adherence was manipulated to be perceived as equal. Again, the effect cannot be attributed to different interpretation of the behavior as evidenced by the absence of effects on behavior evaluation.

The assimilation account of the proposed explanation for the assimilation effect holds that the simple main effect of Group membership on person evaluation in the different adherence condition should be mediated by group evaluations. However, since the assimilation effect is only meaningful compared to the null effect in the equal adherence condition (i.e., in an interaction pattern), the interaction of Group adherence and Target’s group membership on person evaluation should be mediated by group evaluation. This translates into a hypothesis of statistical mediated moderation that Preacher, Rucker, and Hayes (2007) have formalized (Model 2 on p. 194, see also Edwards & Lambert, 2007; Muller, Judd, & Yzerbyt, 2005). This model posits that under different Group adherence conditions (i.e., levels of the moderator) the effect of Target's group membership on the proposed mediator is different, and that the mediator is related to person evaluation. This analysis incorporating the interaction is also a stronger test of the mediation account than two separate simple mediation tests within the Group adherence conditions because it incorporates all cells of the design within the
same analysis and thus efficiently utilizes all available information in the estimation of the error variance for the statistical tests instead of just the data from half of our sample (i.e., the respective Group adherence conditions).

In order to perform the analysis of moderated mediation we created a new variable that holds for each participant the evaluation of the group of which the respective target in the article was a member, henceforth referred to as corresponding group evaluation. For participants in the ingroup target conditions, this variable thus contained the evaluation of the ingroup as a whole and for the remaining participants the evaluation of the outgroup as a whole.

We performed a state of the art analysis of moderated mediation as prescribed by Preacher, Rucker, & Hayes (2007), using the accompanying MODMED macro for SPSS, version 2.2 (Hayes, 2009b). This analysis employs bootstrapping procedures to estimate the indirect effects for each of the values of the hypothesized moderator and confidence intervals for these indirect effects. If the confidence intervals do not overlap, the indirect effects can be considered different and thus it is established that depending on the value of the moderator, there are different indirect effects of the independent on the dependent variable via the mediator. This analysis is complete and sufficient as a statistical test of the moderated mediation hypothesis (Hayes, 2009a; Shrout & Bolger, 2002). We will subsequently complement these concise analyses, for descriptive purposes, by information from separate traditional mediation analyses (Baron & Kenny, 1986) within Group adherence conditions.

The bootstrapping-based analysis (Preacher et al, 2007) based on 5000 resamples revealed, in line with the model of moderated mediation, the indirect effect via group valence in the equal adherence condition was virtually zero, $B_{\text{indirect, equal adherence}} = -.01$,
CI_{95} = [-.15, .12]^2. In contrast, in the different adherence condition the indirect effect was positive, \( B_{\text{indirect, different adherence}} = .38, \ CI_{95} = [.22, .59], \) and, by the logic of parametric significance testing, significantly different from zero as indicated by the fact that the confidence interval does not contain zero. Furthermore, the confidence intervals of the indirect effects do not overlap, indicating that the indirect effects are different to a degree that is equivalent to a .05 \( \alpha \) error level in traditional significance testing. This shows that the \textit{Target's group membership} effect (i.e., the assimilation effect) on person evaluation in the different adherence condition was mediated via \textit{corresponding group evaluation}. In contrast, in the equal adherence condition, where there was no total effect on person evaluation, we also found no indirect effect. A substantial portion of the assimilation effect in person evaluation can thus be explained by differential group valence as a result of the adherence manipulation which then influences person evaluations. The individual paths that are contained in the moderated mediation model are shown in Figure 1 and are taken from two separate simple mediation models (Baron & Kenny, 1986) within each of the \textit{Group adherence} conditions. In this figure, one can also assess whether a BSE remains after the mediator corresponding group evaluation has been partialled out, or in other words: whether a relatively small BSE was overlaid by the assimilation effect. This would result in substantial residual direct effects of \textit{Target's group membership} on person evaluation. However, none of these direct effects are significant (\( \beta = -.14, p = .250 \) and \( \beta = .03, p = .858 \)) and therefore, there is no indication of such a BSE that could have been masked by the assimilation effect.

Discussion

Study 2 replicated the results of Study 1: A difference in adherence to the norm between ingroup and outgroup did lead to an assimilation effect in favor of the ingroup
target. When adherence was manipulated to be equal between the groups, no effect was found. Also, the results of the analyses involving corresponding group evaluations are consistent with the proposed assimilation of person evaluations to the evaluations of the groups they belong to. There was no indication of a BSE occurring simultaneously with the assimilation effect. Such a BSE should have become apparent in the residual, direct effects, once the assimilation effect was controlled for statistically.

General Discussion

Both studies reported here show that violation of a societal norm, which carries stronger negative valence than the norms traditionally used in studies yielding the BSE, does not lead to a BSE. This result may seem to not be surprising at first from a person perception perspective. After all, that judgments about individuals are assimilated towards judgments of a group they belong to, is not novel in itself. However, the assimilation occurs between ingroup and outgroup deviants and the most prominent research on intergroup differences in deviant evaluations strongly suggests the prediction of ingroup derogation. In the past, both if there is no specified different norm adherence between an ingroup and an outgroup as wholes (e. g. Marques et al., 1988, Study 1; Marques et al., 2001, Study 1), as well as if the ingroup as a whole is perceived to hold a superior norm (e. g. Marques et al., 1988, Study 2), downgrading of ingroup deviants has been regularly found. Even if this effect is attenuated by superiority security (Marques et al., 2001, Studies 2 and 3), it is never reversed but eliminated to zero. Thus, while the BSE literature and the SGD may, under specific circumstances, suggest a null effect between ingroup and outgroup person evaluation, it never proposes a reversal of the BSE. We present evidence for such a reversal here: ingroup favoritism in deviant evaluation can occur, namely if a pre-established difference between ingroup
and outgroup carries strong valence to the ingroup's advantage. Then this valence leads to assimilation effects to the advantage of individual ingroup deviants.

Past studies finding BSEs have frequently incorporated a non-deviant condition in which ingroup favoritism was found simultaneously with the ingroup derogation for deviants. We have not used such a condition because the finding of ingroup favoritism for non-deviating group members is very common (Hewstone et al., 2002) and thus, as an isolated result, somewhat unexciting. The theoretical focus of the BSE (as its name suggests) is on norm deviants, not on norm adherents. We are therefore presently not concerned with favoritism in evaluations of targets who adhere to a norm, nor the explanation of such favoritism, but with the meanwhile very fascinating case of intergroup differences in deviant evaluations. In sum, the addition of non-deviant target conditions, while an interesting extension in future research, does not add weight to the basic finding reported here: ingroup favoritism can occur in deviant evaluation instead of a BSE.

Another fruitful extension of the present investigation will be an added different adherence condition in which the outgroup is portrayed as adhering more to the norm in question than the ingroup. According to our assimilation account, ingroup derogation should re-appear in such a condition, but it may be weaker because the general desire for a positive ingroup identity would probably counteract it. Participants may be more reluctant to damage ingroup positivity by downgrading an individual target in addition to the blow to collective positive distinctiveness by the different norm adherence between the groups. Van Prooijen and Lam (2007) have found evidence pointing to that possibility: they reported a BSE only for deviants from a relatively high status group but not in a low status group, indicating that an ingroup must be in a way able to 'afford'
ingroup derogation. The result for higher status groups appears to be at odds with our results. However, the status difference in van Prooijen and Lam (2007) is constituted on a dimension that is very different from the norm being violated and therefore may not be directly applicable to individual target judgments for participants. Such applicability however has been shown to be a necessary condition for assimilation effects in individual target judgments (e.g., Higgins, 1996). In contrast to van Prooijen and Lam's (2007) studies, applicability is obvious in the present experiments as the behavior used to induce group level valence is the same as that of the target. Nevertheless, the integration of the presently reported assimilation effects in connection with reversed intergroup valence difference on the one, and status differences on different dimensions on the other hand is a possible next step in extending the presently reported studies.

It should be noted that, also seemingly contrary to our results, van Prooijen (2006) found a BSE for an unambiguously established violation of a valence carrying norm, the norm not to commit a crime. However, in these studies, differential norm adherence between the groups was not measured or manipulated. In fact, it appears that in these studies, the used ingroup vs. outgroup contexts are especially void of a strong differential adherence perception regarding the norm violation in question (German vs. Dutch football fans illegally selling tickets for a match; red vs. blue department at a fictional factory stealing money at work; students of Free University of Amsterdam vs. Leiden University stealing a bicycle). In contrast, we consistently found a reversed BSE if adherence was explicitly different, but no BSE when adherence did not differ between groups (if anything at all a tendency in that direction emerged in study 1 which is consistent with van Prooijen’s 2006 results).

The present results of course by no means question the BSE or the theoretical
progress associated with its empirical examination. Rather, we would like to contribute to a more detailed picture of group membership effects regarding norm violations. We propose that such effects may be different for different types of norms. Past research in the BSE tradition has already established that violation of a norm is not uniformly subject to ingroup target derogation. Rather, it has been shown that BSEs only appear when the norms are relevant for the perceiver (Marques, 1990, Study 5), they are distinctive between the ingroup vs. the outgroup (Marques et al., 1988, Study 2), or this distinctiveness is sufficiently salient (Marques et al., 1998). Similarly, across several studies, van Prooijen (2006) found evidence that a BSE emerges in punishment recommendations for criminals whose guilt is well established, but a reversal (i.e., an ingroup favoritism effect) for criminals whose guilt is not certain. In sum, a number of boundary conditions have been indentified for the BSE. From our findings on the effects of the generality and valence relevance of the norm in question, we conclude that another process interacts with the processes responsible for the derogation of ingroup deviants: assimilation of individual target judgments to group judgments.

Conclusion

We found, contrary to expectations from past research on intergroup norm deviance, but consistent with our theorizing, that ingroup favoritism emerged if deviants violate a valence-laden norm that is generally adhered to more by the ingroup than the outgroup. This intergroup effect was mediated by differences in group valence due to precisely this different adherence. These findings indicate that the type of norm violated and the perception of collective adherence plays an important role in the perception and evaluation of ingroup vs. outgroup deviants. The resulting pattern of evaluations appears to be more complicated than a global expectation of either ingroup favoritism or
a BSE would suggest. We advocate that future work should continue to both theoretically and empirically disentangle various obviously relevant processes in the evaluation of norm-violating targets regarding such moderations (e.g., relative ingroup status, see van Prooijen & Lam, 2007; uncertainty about behavior authorship, van Prooijen, 2006; opportunity for individual distancing, Eidelman & Biernat, 1999; opportunity for cognitive restructuring of group boundaries, Eidelman et al., 2006). In line with Biernat et al. (1999), we believe that it is desirable to "taxonomize when, where, and how particular processes will 'kick in' and particular patterns of evaluation will emerge" (p. 539, emphasis in original). Much of the research on the BSE has focused on factors producing and strengthening the BSE. Future research may fruitfully attend to situations in which the BSE is weakened or eliminated (van Prooijen & Lam, 2007) or even reversed (as in van Prooijen, 2006, or the currently reported results).

The BSE is a well-documented phenomenon which is especially fascinating in light of the apparent dominant expectation of ingroup favoritism in the evaluation of individual targets. It is an intriguing fact that in situations relatively prone to lead to negative treatment of a target (i.e., norm violations), the target may be at a disadvantage if judged by members of her own group. However, understanding of the phenomenon of ingroup favoritism has certainly profited from research identifying moderating conditions and precluding its indiscriminate extension to other phenomena (e.g., its distinction from explicit outgroup derogation, see Brewer, 1999; Mummendey & Otten, 1998). We think that the BSE deserves equal care in studies on the conditions which attenuate, eliminate or even reverse it.
Footnotes

1 As in study 1, target and behavior evaluations were correlated, \( r(94) = .40, p < .001 \), but were analyzed separately because the patterns of manipulation effects were different.

2 The confidence intervals reported here are corrected for bias (see MacKinnon, Lockwood, & Williams, 2004; Preacher et al., 2007). Uncorrected percentile-based confidence intervals as well as bias-corrected and accelerated confidence intervals (see Preacher et al., 2007; Preacher & Hayes, 2008) yielded the same results.
References


Author Notes

The studies reported are parts of the first author's dissertation completed at the Faculty of Social and Behavioral Science at Friedrich-Schiller-Universität Jena. The reported research was supported by a grant by Deutsche Forschungsgemeinschaft (DFG) awarded to Sabine Otten (DFG OT 170/3). We thank Tobias Raabe, Claudia Schneider, Karoline Hölzer, Kristin Wenzel, and Tina Treml for their help with data collection.
Table 1

*Means (and standard deviations) of person and behavior evaluation as a function of Target group membership and Group adherence in Study 1 (N = 80).*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Target’s group membership</th>
<th>equal</th>
<th>different</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Person evaluation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ingroup</td>
<td>4.90 (0.86)</td>
<td>4.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a (0.90)</td>
<td></td>
</tr>
<tr>
<td>outgroup</td>
<td>4.50 (0.62)</td>
<td>4.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b (0.71)</td>
<td></td>
</tr>
<tr>
<td><strong>Behavior negativity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ingroup</td>
<td>4.26 (1.78)</td>
<td>4.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.19)</td>
<td></td>
</tr>
<tr>
<td>outgroup</td>
<td>4.31 (1.24)</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.98)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Means with different superscripts (column-wise) differ at least $\alpha = .05$ according to adjusted pairwise comparisons (Sidak post hoc comparisons).
Table 2

*Means (and standard deviations) of person and behavior negativity and global ingroup and outgroup evaluation as a function of Target’s group membership and Group adherence in Study 2 (N = 94).*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Target group membership</th>
<th>equal</th>
<th>different</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ingroup</td>
<td>4.44 (0.73)</td>
<td>3.86 (^a) (0.76)</td>
<td></td>
</tr>
<tr>
<td>outgroup</td>
<td>4.20 (0.87)</td>
<td>4.68 (^b) (0.87)</td>
<td></td>
</tr>
<tr>
<td>Behavior evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ingroup</td>
<td>3.84 (1.21)</td>
<td>4.09 (1.14)</td>
<td></td>
</tr>
<tr>
<td>outgroup</td>
<td>3.92 (1.18)</td>
<td>4.17 (1.21)</td>
<td></td>
</tr>
<tr>
<td>Corresponding group evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ingroup</td>
<td>3.41 (0.95)</td>
<td>2.78 (^a) (0.74)</td>
<td></td>
</tr>
<tr>
<td>outgroup</td>
<td>3.38 (0.70)</td>
<td>4.11 (^b) (0.88)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Means with different superscripts (column-wise) differ at at least \(\alpha = .05\) according to adjusted pairwise comparisons (Sidak post hoc comparisons).
Figure Captions

*Figure 1.* Illustration of the different indirect mediation effects between *Target group membership* and person evaluation via group valence. Separate models for a) equal adherence and b) different adherence of the groups conditions are shown. Individual paths are estimated using traditional regression analyses as prescribed by Baron and Kenny (1986).
Figure 1

a) Mediation for equal adherence condition

Group evaluation

$\beta = .02, p = .917$

Target's group membership
(1 = ingroup, -1 = outgroup)

$\beta = -.14, p = .250$

($\beta = -.15, p = .314$)

Person evaluation

Sobel test: $z = -0.11, p = .914$

b) Mediation for different adherence condition

Group evaluation

$\beta = .68, p < .001$

Target's group membership
(1 = ingroup, -1 = outgroup)

$\beta = .03, p = .858$

($\beta = .46, p = .001$)

Person evaluation

Sobel test: $z = 4.15, p < .001$