EASP WILEY

'(N)One of us but all of them!' Ingroup favouritism on individual and group levels in the context of deviant behaviour

Zahra Khosrowtaj[1] | Sarah Teige-Mocigemba[1] | Vincent Yzerbyt[2]

[1]Department of Psychology, University of Marburg, Marburg, Germany

[2]Faculty of Psychology and Educational Sciences, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Correspondence
Zahra Khosrowtaj, University of Marburg, Marburg, Germany.
Email: zahra.khosrowtaj@uni-marburg.de

Abstract

Past research hints both at more extreme judgements of ingroup deviants and at attributional biases in the case of Muslims, immigrants and refugees. We examined two recently observed patterns in the context of intergroup violence: harsher judgements on the individual level (black sheep effect) and milder judgements on the cultural level when a perpetrator stems from the ingroup. We further investigated whether these patterns were affected by (a) the outgroup being salient (Experiment 1), (b) the comparison context (Experiments 2–3) and (c) participants perceiving the ingroup as high versus low in entitativity (Experiment 3). Experiments 1 ($N = 437$), 2 ($N = 283$) and 3 ($N = 703$) revealed the presence of robust effects on cultural level with participants treating the ingroup culture more leniently than the outgroup culture. Moreover, on the individual level, Experiments 2 and 3 found an overall black sheep effect that was especially prevalent in an intergroup context. Outgroup salience and ingroup entitativity did not affect participants' judgements on individual and cultural levels. This protection of the ingroup both on an individual and on a cultural level may hint at a derogation of the outgroup. We discuss implications and insights for future research.

KEYWORDS
Black sheep effect, culture blaming, discrimination, entitativity, ingroup protection, intergroup attribution, outgroup derogation

1 | INTRODUCTION

(…) when I, who am visibly Muslim, cross the street at a red light, 1.9 billion Muslims are crossing the street with me. (Gümüsay, 2022, p. 57)

I never hear: You are totally intelligent because you were born in Iran or in Palestine or in Israel. But: You are misogynistic because you were not born in Germany. (Fereidooni interviewed by Seelig, 2023)

In Berlin (Germany), the police arrested suspects from 18 nationalities during the night of New Year's Eve 2022/2023 (Tagesschau, 2023a, 2023b; Windisch et al., 2023). Among those, 45 were of German nationality, followed by 27 Afghans and 21 from Syria (Tagesschau, 2023b; Windisch et al., 2023). The incidents of that evening triggered waves of debates regarding a failed integration. The Christian Democratic Union (CDU), a major German political party, asked for the forenames of those Germans referring to those with a dual nationality (Moll, 2023). The leader of the CDU, Friedrich Merz, attributed the crimes to the Arabic background of the perpetrators (Becker, 2023), pointing to a problematic culture, as seen elsewhere (e.g., Stürmer et al., 2019). Ingroup perpetrators may seem to be exceptions to the

rule (Carnaghi & Yzerbyt, 2007; Kunda & Oleson, 1995; Parks-Stamm, 2013), while outgroup perpetrators seem to belong to a homogenous and indeed threatening culture. The present endeavour explores two strategies that group members may use to protect the ingroup, namely, perceiving an ingroup deviant as one bad apple while seeing the culture of an outgroup deviant as an underlying cause. We examined these two strategies in the context of intra- and intergroup situations that involve Germans as ingroups and Muslims[1] as outgroup.

1.1 | Black Sheep Effect and Culture Blame

Previous research has shown that we tend to blame deviant ingroup members more extremely than comparable outgroup members to maintain the positivity of the whole ingroup (Abrams et al., 2000, 2002; Marques, 1990; Marques et al., 1988; Marques & Yzerbyt, 1988; Yzerbyt et al., 2000). For instance, students judged poor speeches by ingroup members more negatively than poor speeches by outgroup members (Marques & Yzerbyt, 1988). This so-called black sheep effect has been conceptualized as a sophisticated form of ingroup favouritism because the deviant member would seem to no longer come across as a 'standard' ingroup member (Marques et al., 1988; Marques & Yzerbyt, 1988).

Extending this tradition of work, recent work hints at the possibility that ingroup protection could be served both on individual and cultural levels (Khosrowtaj et al., 2024c). In this study, participants worked through an alleged newspaper article and judged either an ingroup or an outgroup perpetrator. Results confirmed the presence of a black sheep effect. Moreover, participants evaluated the ingroup culture more leniently than the outgroup culture (Khosrowtaj et al., 2024c). This initial evidence hints at a potentially important extension of the original work on the black sheep effect. Indeed, the protection of the ingroup has heretofore been conceived only in terms of the evaluation of the individual deviant targets. However, it may be that evaluations also go beyond the individual deviant target and address the groups in presence. In the present work, we wanted to examine the robustness of this initial finding and examine the manifestation of ingroup protection both at the individual level (i.e., the black sheep effect) and the level of the group or, more generally, the culture.

As it turns out, this pattern of group judgements is reminiscent of the intergroup attribution bias and the ultimate attribution error. A negative behaviour of a fellow ingroup member comes across as an exception to the rule or as due to causes beyond individual control. In contrast, the same negative behaviour performed by an outgroup member is attributed to the deep characteristics of this person or of his group as a whole (Duncan, 1976; for a review, see Hewstone, 1990; Islam & Hewstone, 1993; Pettigrew, 1979; Taylor & Jaggi, 1974). In a classic illustration relying on an interracial context, a physically violent

behaviour perpetrated by a black (vs. white) person was attributed to personal (vs. situational) factors (Duncan, 1976). Building on this line of work, the present series of experiments replicates and extends Khosrowtaj et al.'s (2024c) findings. Our aim is thus to shed further light on people's comparative judgement when they witness a deviant act, and we do so by looking at the evaluation of the individual ingroup or outgroup perpetrator and his culture. As mentioned earlier, only one experiment so far assessed the evaluations of participants on the cultural level. The ambition of the present replication was to test whether the pattern obtained in this initial study constitutes a robust finding. A replication of the obtained pattern would lend further credence to the need to extend the original black sheep effect and show that ingroup protection also takes place at a cultural level. In the present experiments, we predicted ingroup favouritism in the form of the black sheep effect at the individual level and leniency towards the ingroup on the cultural level. Specifically, we did so by building on the ongoing debates regarding the societal threats posed by Muslims and the Muslim culture in Western countries.

1.2 | Muslims as a Societal Threat

Several lines of work provide evidence of the existing hostility towards Muslims in the Western world. Indeed, there are numerous reports of resistance to Muslim immigration (Gusciute et al., 2021; Liebe et al., 2018), the decreased acceptance of Arab or African immigrants (Czymara & Schmidt-Catran, 2017), the impact of terrorist attacks in the name of Islam on attitudes towards immigrants (Ferrin et al., 2020), the aggressive behaviour towards Muslim-appearing targets as part of a shooter paradigm (Unkelbach et al., 2008) or even the discrimination in hiring contexts (Di Stasio et al., 2019; Unkelbach et al., 2010). Perceiving Muslims as threatening has been shown to predict anti-Muslim policies as well as the willingness to participate in extremist behaviour targeting Muslims (Dunwoody & McFarland, 2018).

Media analyses also reveal the presence of a representational bias against Muslim and foreign perpetrators (Hestermann, 2019; Kearns et al., 2019) and attributional biases (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022), which include direct attributions of deviant acts to the outgroup background. Inspired by these media analyses, previous work investigated whether this bias already had an impact on information search (Khosrowtaj et al., 2024a). Specifically, participants imagined reading a newspaper article depicting a crime before indicating their interest in a list of (non)-stereotypic information categories that they wished to know more about. In line with the representational bias, participants indicated higher interest in stereotypic information categories such as religious affiliation when confronted with a perpetrator coming from a predominantly Muslim country than when facing a German perpetrator (Khosrowtaj et al., 2024a). Among other findings (e.g., Fitzgerald et al., 2012; Hirtenlehner, 2019), a number of empirical studies confirm that Muslims or Muslim-appearing targets are associated with violence and crimes in Western countries (Unkelbach et al., 2008, 2010).

---

[1] Even though we operationalized the group membership of the perpetrators by their nationality, there is a racialization of Muslims and a synonymous use of religious affiliation of the outgroup and ethnicity or nationality attributions, although a religious affiliation does not translate to a nationality nor a nationality does translate to a religious affiliation (Kteily et al., 2015; Shooman, 2012).

Given our present aim to investigate ingroup protection on individual and group levels, we decided to rely on this specific societal context.

1.3 | The Present Experiments

Although the black sheep effect has so far been investigated at the level of the individual member (Marques et al., 1988; Marques & Yzerbyt, 1988), the present experiments build upon the observation that physical violence is often being attributed to an alleged Islamic culture (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022).

Specifically, we conducted three experiments to replicate and extend the research showing the protection of the ingroup both on the individual level (i.e., more ingroup perpetrator blaming than outgroup perpetrator blaming) and on the cultural level, that is, less ingroup culture blaming than outgroup culture blaming (Khosrowtaj et al., 2024c). In all experiments, we used alleged newspaper articles (Experiment 1) or newsflashes (Experiments 2–3) where a perpetrator is accused of rape. We manipulated group membership using different names (cf. Kauff, 2022, for a similar approach using names for categorization) and nationalities (Khosrowtaj et al., 2024c). We then assessed participants' judgements on an individual and cultural level. We predicted a black sheep effect on individual blame measures (detailed below) while on a cultural level, we expected to observe a protection of the ingroup culture. To secure an ingroup white sample, we relied on participants with German as their first language across all experiments and excluded those with parents with a mother tongue from a predominantly Muslim country. Across all experiments, we collected data using Prolific Academic, ensuring that participants who took part in the previous experiments of this line of work would be excluded beforehand.

We also wanted to examine some boundary conditions of our predicted effects. Experiment 1 checked whether making the outgroup salient before participants learned about the crime resulted in a more pronounced black sheep effect and ingroup culture protection. In Experiments 2–3, we considered the role of the comparison context using a within-participant design with either the ingroup or the outgroup newsflash being presented first. Experiment 3 further tested the impact of ingroup entitativity (Campbell, 1958; Lewis & Sherman, 2010) on the judgements at both individual and cultural levels. We pre-registered all the projects and the data as well as the analysis syntax for all three experiments are available at OSF https://osf.io/3mgj5/. This manuscript and its supplementary materials include all studies, measures, manipulations and participant exclusions. We collected some sociodemographic variables for exploratory purposes for a follow-up project. As suggested by the editor, the confidence intervals for the effect sizes across all experiments are reported as part of the supplementary materials.

2 | EXPERIMENT 1

The goal of Experiment 1 was to test ingroup protection at the individual and the cultural level. In line with a host of previous demonstrations

of the black sheep effect, we informed participants about an event that involved a perpetrator and a victim who belonged to the same group, either the ingroup or the outgroup. This strategy allowed examining judgements in a context that is devoid as much as possible of potential demand effects regarding intergroup comparison, especially in the case of the ingroup situation.

To the extent that the black sheep effect emerges as a response to an endangered ingroup's image (Yzerbyt et al., 2000), the (symbolic) presence of a threatening outgroup may reinforce the urge for intergroup differentiation (cf. Brewer, 1991; Haslam & Oakes, 1995). Indeed, previous work indicates that thinking of Arabs and Muslims (rather than no category) influenced participants' responses in a shooter paradigm task (Mange et al., 2012). That is, the mere thought of threatening categories may suffice to exert an influence on participants' judgements and behaviour. In line with these findings, we wanted to examine the impact of mere outgroup salience (cf. Yuki & Yokota, 2009) by means of a priming manipulation on the judgements about ingroup and outgroup perpetrators as well as their cultures.

In sum, Experiment 1 adopted a 2 (couple: ingroup vs. outgroup) × 2 (priming manipulation: outgroup salient vs. outgroup not salient) factorial design. We predicted a black sheep effect at the individual level (perpetrator blaming and verdict). We further expected a similar pattern for the classification of the incident as rape with participants perceiving the behaviour more as rape in the case of an ingroup than an outgroup perpetrator. As for the cultural level, we predicted the exoneration of the ingroup compared to the outgroup culture.

Regarding the priming manipulation, we expected outgroup salience to influence the ingroup couple condition more than the outgroup couple condition. Because making the outgroup salient would place participants in the ingroup couple condition in an intergroup rather than an intragroup context (cf. Haslam et al., 1995; Haslam & Oakes, 1995), we expected a harsher black sheep effect compared to the one observed in the no priming condition where no such intergroup context was created. In other words, the mere salience of the outgroup (priming condition) should lead participants to manifest a stronger black sheep effect than in the no priming condition. This is because participants may compare the depicted ingroup perpetrator with the threatening outgroup (intergroup context) and, accordingly, may express more severe judgements towards the ingroup perpetrator as a distancing strategy with respect to one negative exemplar. We also expected outgroup salience to influence the classification of the perpetrator's behaviour as rape in the ingroup couple condition by increasing the difference between the classification as rape for the ingroup compared to the outgroup perpetrator. In a similar vein, we expected outgroup salience to be conducive to more ingroup culture exoneration.

2.1 | Method

2.1.1 | Sample size estimation

We estimated sample size using G*Power 3.1.9.2 for linear multiple regression: Fixed model, R^2 deviation from zero with 7 predictors, $\alpha = .05$, power = .95 and an estimated small effect size $f^2 = .05$. This

analysis resulted in a total sample size of $N = 444$. Taking possible dropouts into account, we aimed for a sample size of $N = 500$.

We used G*Power 3.1 to conduct a sensitivity power analysis (linear multiple regression: fixed models, R^2 deviation from zero) using the final sample size ($N = 437$). For a power of .95, $\alpha = .05$ with 3 predictors, the minimum effect that could be obtained was $f^2 = .039$.

2.1.2 | Participants

A total of $N = 443$ participants from the Prolific Academic platform (www.prolific.co) took part in two-time points of a larger project including Experiment 1, which was assessed at T2. For other work, we collected different self-report measures at T1 (time interval 14–23 days) and invited the same participants to T2 (present Experiment). Demographics (e.g., sex, age and mother tongue) were assessed at T1.

As was pre-registered, in Experiment 1, we excluded participants if they (a) did not have German as mother tongue, (b) had parents with a mother tongue from a predominantly Muslim/Arab country, (c) had extreme values in the time they spent at the manipulation page (priming) indicating that they did not follow instructions and (d) had extreme values in the time they spent between the manipulation (priming) and the vignette (indicating that possible effects of the experimental manipulation on the DVs could have faded). Moreover, and prior to data analyses, we decided to exclude participants if (e) the studentized deleted residuals for the dependent measures were $> \pm 4$ as recommended by Judd et al. (2011). Note that participants with a mother tongue from a predominantly Arab/Muslim country were not invited to the main Experiment (T2). In light of these exclusion criteria, we excluded six participants, all of which had a dwell time of 3 SD above the mean time spent between the priming manipulation and the vignette ($M = 14.43$ ms; $SD = 24.90$). Thus, the following analyses relied on a total sample of $N = 437$[2] ($n = 160$ female, $n = 224$ male, $n = 2$ diverse, $n = 51$ did not indicate their gender) with an average age of $M = 31.16$ years ($SD = 10.41$; range: 18–68 years; $n = 6$ participants did not indicate their age).

2.1.3 | Procedure and measures

Overall procedure. Participants worked through two alleged separate studies. The first part involved the priming manipulation (outgroup salient vs. outgroup not salient) and the second part included a fictitious newspaper article (see below) and the assessment of blame and verdict judgements (see below). We randomly assigned participants to one of the priming conditions. A second randomization followed regarding the vignette with either an ingroup or an outgroup couple. To avoid possible confounds, we further asked participants to indicate if they had been victims of sexual violence (sexual abuse

$n = 38$, forced intercourse, $n = 24$, almost forced to engage in sexual acts, $n = 40$).[3] Finally, participants had the possibility to contribute open remarks regarding the studies before receiving a full debriefing.

Priming manipulation. The experimental manipulation included a bogus quiz including questions with open response fields. Participants learned that the aim of Study 1 (priming manipulation) was the selection of appropriate quiz questions for a new general knowledge test. In the outgroup not salient condition, participants read 10 general quiz questions, which did not refer to an in- or outgroup (e.g., 'How many bones does the human body have?', 'How many keys does the piano have?'). In the outgroup salient condition, 5 of the 10 questions made the outgroup salient (i.e., 'How many Muslims live in Germany?' and 'How many refugees live in Germany?').

Newspaper article. We adopted the fictitious newspaper article from previous work (Khosrowtaj et al., 2024c), but only included couples with the same ethnic background. The newspaper article read as follows:

> Frankfurt am Main
>
> **Suspected rape:**
> **Police interrogate Ahmed S. (Andreas S.)**
>
> 21. Mai 2018, 11:32 am.
>
> According to the police report, 28-year-old Latifa K. (Lena K.). was raped last Friday. The student met the suspect Ahmed S. (Andreas S.) through a dating app in early April. A meeting took place at the man's home last week. Small amounts of alcohol were consumed.
>
> When the young woman missed the last S-Bahn, the accused is said to have offered her to sleep on the couch. The 35-year-old is said to have raped the Afghan (German) woman that night. Latifa K. (Lena K.) contacted the police the following day.
>
> When the accused Afghan (German) Ahmed S. (Andreas S.) was picked up by the police for questioning, he was visibly surprised and denied the allegations of rape. Although there were sexual acts with Latifa K. (Lena K.) in his apartment, they are said to have taken place consensually.

Judgements. Based on previous work, we assessed perpetrator (Cronbach's $\alpha = .84$) and victim blaming (Cronbach's $\alpha = .89$) with the following six items: 'How much is (name of the ingroup/outgroup member) to be held responsible for what has occurred?', 'How much influence did (name of the ingroup/outgroup member) have on the

---

[2] Note that the results were generally consistent with the results of analyses without any exclusions (for details, please see the Supplementary Materials).

[3] Leaving out all participants with a history of sexual violence yielded a total $N = 388$ remaining for the analysis. Running the analyses without these participants did not change the reported result patterns. Thus, participants who experienced being a victim themselves remained in the analysis sample.

outcome of the situation?' on a 7-point scale varying from 1 (*not at all responsible*) to 7 (*fully responsible*) (Süssenbach et al., 2012, 2017), 'How likely do you think it is that (name of the ingroup/outgroup member) could have avoided the incident?', 'How much do you think (name of the ingroup/outgroup member) had control over the situation?' on a 7-point scale varying from 1 (*not at all*) to 7 (*very much*) (Bieneck & Krahé, 2011), 'How angry do you feel toward (name of the ingroup/outgroup member)?' and 'How hostile do you feel toward (name of the ingroup/outgroup member)?' on a 7-point scale varying from 1 (*no anger/hostility at all*) to 7 (*very much anger/hostility*) (van Prooijen, 2006). For the sake of the scope of this work, we will focus on the perpetrator in the following.

Next, we measured the verdict given to the perpetrator with three items (Cronbach's α = .85): 'How likely is (name of the ingroup/outgroup member) guilty of a crime?' (Süssenbach et al., 2012, 2017) and 'How strongly should (name of the ingroup/outgroup member) ought to be held criminally liable for the crime?', both rated on a 7-point scale from 1 (*not at all*) to 7 (*very much*) (Bieneck & Krahé, 2011), and 'What sentence length do you consider appropriate?' rated on a scale from 1 (*acquittal*) to 7 (*6 years or more*) (Süssenbach et al., 2012, 2017). Participants then indicated the extent to which they classified the behaviour of the perpetrator as rape (Bridges, 1991) on a 7-point scale ranging from 1 (*definitely not rape*) to 7 (*definitely rape*).

Then, building on previous efforts (Khosrowtaj et al., 2024c), we asked participants to indicate the blame that they attributed to the culture of the perpetrator (4 items, Cronbach's alpha = .94) and the victim (4 items, Cronbach's alpha = .96). The items read as follows: '(Name)'s behaviour can in part be explained by cultural norms and values', 'How much influence did the (name of the ingroup/outgroup member) culture have on the outcome of the situation?', 'How much is (name of the ingroup/outgroup member) culture to be held responsible for what has occurred?', '(Name)'s culture is partly to blame for the events'.' on 7-point scales ranging from 1(*no influence at all*) to 7 (*very strong influence*). Finally, as part of a different project and beyond the scope of the present work, we also asked participants to answer four items targeting the typicality of the perpetrator and the victim as well as their own similarity to the perpetrator and victim (cf. Bettencourt et al., 1997).

2.2 | Results

We tested our predictions with a series of multiple regression models using contrast codes for the factors couple (ingroup = −0.5 and outgroup = 0.5) and priming (control = −0.5, priming = 0.5). We included the interaction term between couple and priming in all analyses.

2.2.1 | Individual blame

Perpetrator blaming. The analysis with perpetrator blaming as a criterion revealed no significant main or interaction effects, all *p* values ≥ .553 (see Figure 1a).

Verdict. The analysis with verdict as criterion showed no significant main or interaction effect, all *p* values ≥ .491 (see Figure 1b).

Classification as rape. The analysis with the classification of the incident as rape revealed no significant effect of couple nor priming manipulation, all *p* values ≥ .548 (see Figure 1c).

Culture Blaming. The analysis with the perpetrator's culture blaming as a criterion indicated the predicted main effect of the couple, b = 0.37, $t(433)$ = 2.29, p = .023, η_p^2 = .012, 95% CI [0.00–0.04] such that participants blamed the ingroup culture less (M = 2.89, SD = 1.56) than the outgroup culture (M = 3.27, SD = 1.77). There was no other main or interaction effect, all *p* values ≥ .520 (see Figure 1, panel d)[4].

2.3 | Discussion

In contrast to previous work (Khosrowtaj et al., 2024c), we did not find a black sheep effect, whether on perpetrator blame, verdict or classification of the perpetrator's behaviour as rape. One explanation for this pattern may reside in the ambiguity of the newspaper article with respect to the guilt of the perpetrator. Indeed, it has been shown that when an ingroup's guilt is certain, the ingroup deviant is judged more harshly compared to an outgroup counterpart. In contrast, when there is a so-called benefit of the doubt (that is, the ingroup guilt is not certain) discrimination against an outgroup member is more likely to emerge (Otten & Gordijn, 2014; van Prooijen, 2006; but see, Khosrowtaj et al., 2024b). Recall that the newspaper articles of Experiment 1 described the perpetrator to be surprised about the deviant claim. This kind of information may have raised doubts about guilt and prevented the emergence of harsher judgments of the ingroup perpetrator. As expected, however, we observed a general leniency towards the ingroup culture compared to the outgroup culture.

Our priming manipulation did not have the predicted effect on perpetrator and cultural judgements. The five outgroup questions that were embedded in an alleged quiz did likely not prove sufficient to make the outgroup salient. Put differently, the priming manipulation included in the alleged first study failed to activate stereotypes about the outgroup in a way that would carry over to the main dependent variables measured in the alleged second study. Furthermore, the outgroup questions we used (e.g., 'How many Muslims live in Germany?') may have been too broad to create an intergroup context (Haslam et al., 1995; Haslam & Oakes, 1995) as a necessary boundary condition for the black sheep effect to occur. Indeed, it has been shown that primes induce assimilation or contrast depending on their breadth: broad ingroup primes and narrow outgroup primes provided the greatest assimilation and contrast (agreement and disagreement with one's political in- and outgroup), respectively (Ledgerwood & Chaiken, 2007). Accordingly, too broad outgroup questions may prevent contrasting effects on the perpetrator and culture judgements.

---

[4] All conclusions remain unchanged when we reran the analyses by including the participants with a dwell time +3SD.
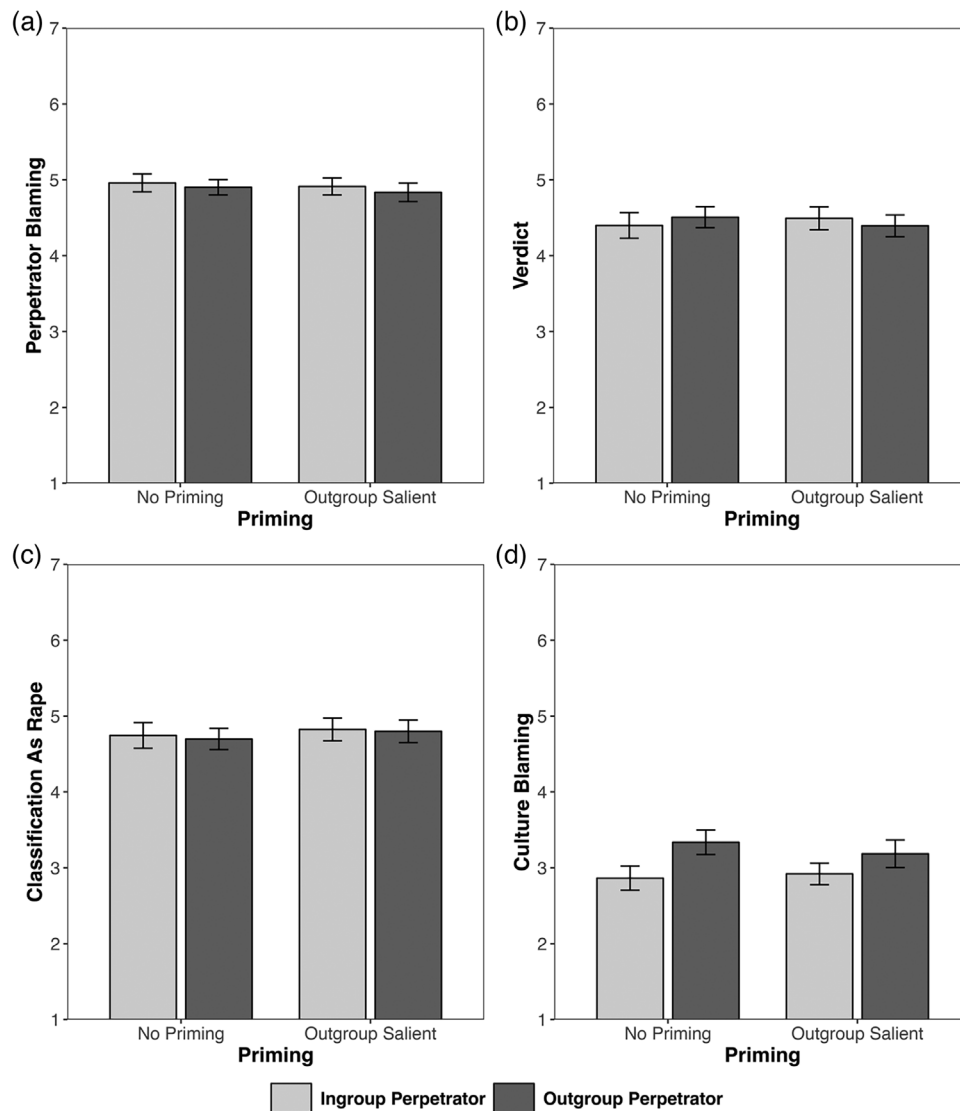
FIGURE 1 *Judgements as a function of couple and priming manipulation. Notes.* a: Perpetrator blaming as a function of couple and priming manipulation. b: Verdict as a function of couple and priming manipulation. c: Classification as rape as a function of couple and priming manipulation. d: Culture blaming as a function of couple and priming manipulation. Higher scores indicate harsher judgments. Error bars indicate the standard error of the mean.

Considering these limitations, we decided to conduct another experiment to test ingroup protection at the individual and the group level in a design that we hoped would maximize the chances for a black sheep effect to emerge. First, we created newsflashes that did not leave room for any ambiguity about the guilt of the perpetrator. Second, we secured a clear intergroup context by using a within-participants manipulation of the group membership of the perpetrator. That is, we provided participants with two newsflashes targeting perpetrators, one being from the ingroup and the other from the outgroup. Such a design should increase the likelihood that participants appraise the perpetrators in the context of an intergroup comparison, especially in the case where the ingroup perpetrator is presented second.

3 | EXPERIMENT 2

Experiment 2 used a 2 (perpetrator: ingroup vs. outgroup) × 2 (order: ingroup first vs. outgroup first) mixed design with the first factor varying within participants and the second between them. Participants worked through two newsflashes (in counterbalanced order) while having two settings (dating vs. cinema) for increased credibility of the two newsflashes.

As the newsflashes were less ambiguous about the perpetrator's guilt and due to the salient intergroup context (within-participant design), we predicted a black sheep effect with harsher judgements of the ingroup than of the outgroup perpetrator on all measures targeting the individual: perpetrator blaming, verdict and classification of

perpetrator's behaviour as rape. We further expected to replicate the exoneration of the ingroup culture rather than the outgroup culture. Finally, we also aimed to examine whether the order of presentation affects the judgements. If any, order effects may arise due to the outgroup first condition where the intergroup context is salient (e.g., Haslam & Oakes, 1995).

3.1 | Method

3.1.1 | Sample size estimation

We ran a power analysis using the PANGEA (v0.2) web app (https://jakewestfall.shinyapps.io/pangea/) to estimate the sample size required for observing a small effect for the main effect of culture blaming. For our 2 (order) × 2 (perpetrator) factorial design participants were nested in order and crossed with the perpetrator. Power analysis was based on the main effect of the factor perpetrator in Experiment 1 for culture blaming. Here, we estimated a $d = 0.24$. This resulted in $N = 300$ for achieving a power of 94.9%.

3.1.2 | Participants

A total of $N = 303$ participants from the Prolific Academic (www.prolific.co) completed the online experiment. As was pre-registered, in Experiment 2, we excluded participants if they (a) did not have German as their mother tongue, (b) had parents with a mother tongue from a predominantly Muslim/Arab country and (c) participated at T2 of Experiment 1. Also, prior to data analyses, we decided to exclude participants if they (d) indicated having participated in a previous similar experiment at the University of Marburg, (e) spent too little time on the vignette pages (≤ 5 seconds) and (f) the studentized deleted residuals for the dependent measures were $> \pm 4$ as recommended by Judd et al. (2011). The two latter exclusion criteria were not pre-registered but added prior to data analyses for making sure that participants followed instructions and paid attention to the tasks while working through the experiment. Following these exclusion criteria, we excluded 20 participants, $n = 7$ with parents with a mother tongue from a predominantly Muslim country; $n = 1$ who participated in a previous similar experiment; $n = 3$ with too short dwell times (≤ 5 seconds) on the pages showing the vignettes and $n = 9$ based on the studentized deleted residuals with absolute values > 4 (Judd et al., 2011).[5] Our final analyses relied on a total sample of $N = 283$[6] ($n = 121$ female, $n = 154$ male, $n = 4$ diverse and $n = 4$ did not indicate their gender) with an average age of $M = 28.41$ years ($SD = 7.61$; range: 18 – 66 years).[7]

---

[5] Regarding this criterion, no dependent measures collected for Experiment 1 revealed values outside the range −4 to +4.

[6] Note that the results were generally consistent with the results of analyses without any exclusions (for details, please see the Supplementary Materials).

[7] One person indicated an age of 7 and was excluded only for assessing these descriptive statistics. However, this participant remained in the final sample as participants from Prolific Academic must be at minimum 18 years old and as this subject did not apply as conspicuous due to the dwell times.

We conducted a sensitivity power analysis (statistical test: analyses of variance (ANOVA), fixed effects, main effects and interactions) using the final sample size ($N = 283$). For a power of .95 $\alpha = .05$, numerator $df = 1$ and number of groups = 2 (factor perpetrator), number of covariates = 0, the minimum effect size that could be obtained was $f^2 = .215$.

3.1.3 | Procedure and measures

Overall procedure. Participants worked through a study called 'perception of sexual assaults', which we framed as an international study examining the perception and judgement of sexual assaults in different cultures. We assigned participants randomly to the experimental condition. All participants worked through two newsflashes starting either with the ingroup or the outgroup.

Newsflashes. We created two different contexts for the newsflashes which read as follows:

> *Hannover*
>
> **Rapist caught**
>
> 21.05.2018, 11:32 am
>
> The 35-year-old Afghan (German) Ahmed S. (Andreas S.) is accused of having committed rape last Friday. The contact was initiated via a dating app. Ahmed S. (Andreas S.) invited to dinner in his apartment. Later that night—despite resistance—the afore-said rape occurred. According to the police report, he initially denied the crime, but after being confronted with the clear means of evidence, he confessed to the rape.

> *Kiel*
>
> **Rapist caught during the crime**
>
> 11.12.2010, 10:30 pm
>
> The German (Afghan) 28-year-old Andreas S. (Ahmed S.) is reported for having committed rape after visiting a cinema. At first, he initiated his approach in a friendly way—similar to a harmless flirt. However, after his interest was not returned, he became aggressive and violent. Passers-by surprised the perpetrator not far from the cinema still in the act and prevented his escape. Andreas S. (Ahmed S.) finally confessed to the crime after trying to convince by claiming consensual acts.

Note that we avoided mentioning the victim explicitly so as to rule out possible identification processes with the victim (e.g., see George & Martínez, 2002, on the judgment of inter- and intragroup perpetrators and victims; but see Khosrowtaj et al., 2024c) and for ruling out any other inter- or intragroup context beyond the one created by the within-subjects design.

Judgements. After reading each newsflash, participants worked through the perpetrator blaming measure (6 items; Cronbach's alpha$_{ingroup}$ = .60, Cronbach's alpha$_{outgroup}$ = .64), verdict measure (3 items; Cronbach's alpha$_{ingroup}$ = .64, Cronbach's alpha$_{outgroup}$ = .72), the single item targeting the classification of the perpetrators' behaviour as rape and the perpetrator's culture-blaming measure (4 items; Cronbach's alpha$_{ingroup}$ = .94, Cronbach's alpha$_{outgroup}$ = .96), see Experiment 1 for exemplary items. Finally, participants provided demographics and were thanked for their participation.

3.2 | Results

To test our predictions, we ran several 2 (perpetrator: ingroup vs. outgroup) × 2 (order: ingroup first vs. outgroup first) mixed-model ANOVAs.

3.2.1 | Individual blame

Perpetrator blaming. The analysis with perpetrator blaming as a criterion revealed a significant main effect of the perpetrator, $F(1,281) = 5.31$, $p = .022$, $\eta_p^2 = .02$, 95% CI [0.00–0.06], indicating that participants blamed the ingroup perpetrator more strongly ($M = 6.28$, $SD = 0.69$) than his outgroup counterpart ($M = 6.21$, $SD = 0.73$). We further observed a marginally significant effect of order, $F(1,281) = 3.69$, $p = .056$, $\eta_p^2 = .01$, 95% CI [0.00–0.05], such that the overall mean ratings were harsher when the ingroup was presented first ($M = 6.32$, $SD = 0.65$) than when the outgroup was presented first ($M = 6.17$, $SD = 0.77$). We further observed a significant interaction, $F(1,281) = 39.35$, $p < .001$, $\eta_p^2 = .12$ and 95% CI [0.06–0.20] (see Figure 2, panel a). To probe this interaction, we examined simple effects for the perpetrator, separately for each order. When the outgroup newsflash was presented first, the ingroup perpetrator was blamed more ($M = 6.30$, $SD = 0.73$) than the outgroup perpetrator ($M = 6.04$, $SD = 0.79$), $t(281) = -6.01$, $p < .001$, $\eta_p^2 = .11$ and 95% CI [0.05–0.19]. When the ingroup newsflash came first, the outgroup perpetrator was blamed more ($M = 6.38$, $SD = 0.63$) compared to his ingroup counterpart ($M = 6.26$, $SD = 0.65$), $t(281) = 2.83$, $p = .005$, $\eta_p^2 = .03$, 95% CI [0.00–0.08].[8]

Verdict. The analysis with the verdict as a criterion again revealed a significant main effect of the perpetrator, $F(1,281) = 4.32$, $p = .038$, $\eta_p^2 = .02$ and 95% CI [0.00–0.06], indicating that participants blamed the ingroup perpetrator more strongly ($M = 6.28$, $SD = 0.83$) than his outgroup counterpart ($M = 6.21$, $SD = 0.88$). We further observed a significant main effect of order $F(1,281) = 5.00$, $p = .026$, $\eta_p^2 = .02$ and 95% CI [0.00–0.06], such that the overall mean ratings were harsher when the ingroup was presented first ($M = 6.35$, $SD = 0.80$) than when

the outgroup was presented first ($M = 6.14$, $SD = 0.91$). We further observed a significant interaction, $F(1,281) = 63.83$, $p < .001$, $\eta_p^2 = .19$ and 95% CI [0.11–0.26] (see Figure 2, panel b). We again examine the simple effects of perpetrators, separately for each order. When the outgroup newsflash was presented first, the ingroup perpetrator was blamed more ($M = 6.31$, $SD = .81$) than the outgroup perpetrator ($M = 5.96$, $SD = 0.96$), $t(281) = -7.06$, $p < .001$, $\eta_p^2 = .15$ and 95% CI [0.08–0.23]. When the ingroup newsflash was presented first, the outgroup perpetrator was blamed more ($M = 6.45$, $SD = 0.73$) than his ingroup counterpart ($M = 6.25$, $SD = 0.85$), $t(281) = 4.22$, $p < .001$, $\eta_p^2 = .06$ and 95% CI [.02; .12].[9]

Classification as rape. The analysis of the classification of the incident as rape revealed a significant perpetrator × order interaction, $F(1,281) = 21.59$, $p < .001$, $\eta_p^2 = .07$ and 95% CI [0.02–0.14] (see Figure 2, panel c). There was no other significant effect, all p's > .387. Turning to simple effects of perpetrator for each order separately revealed that, when the outgroup newsflash was presented first, the ingroup perpetrator's behaviour was more likely classified as rape ($M = 6.77$, $SD = 0.55$) than the outgroup perpetrator's behaviour ($M = 6.50$, $SD = 0.84$), $t(281) = -3.86$, $p < .001$, $\eta_p^2 = .05$ and 95% CI [0.01–0.11]. When the ingroup newsflash was presented first, the outgroup perpetrator's behaviour was more likely classified as rape ($M = 6.79$, $SD = 0.58$) than the behaviour of his ingroup counterpart ($M = 6.60$, $SD = 0.83$), $t(281) = 2.69$, $p = .007$, $\eta_p^2 = .03$ and 95% CI [0.00–0.07].[10]

Culture blaming. Turning to culture blaming, the predicted main effect of perpetrator proved significant, $F(1,281) = 116.22$, $p < .001$, $\eta_p^2 = .29$ and 95% CI [0.21–0.37] such that the ingroup culture came across as less responsible for the deviant act ($M = 2.42$, $SD = 1.28$) than the outgroup culture ($M = 3.15$, $SD = 1.58$). There was no other effect, all p values ≥ .087 (see Figure 2, panel d).[11]

3.3 | Discussion

In Experiment 2, we observed the predicted overall black sheep effect for two out of the three individual blame measures, namely, perpetrator blaming and verdict. We further replicated the protection of the ingroup on cultural terms. Interestingly, the significant interaction indicates that participants manifested a black sheep effect only when the outgroup newsflash was presented first. This is an important finding because it fits entirely with our rationale, underlying the priming manipulation of Experiment 1. Indeed, when the outgroup is

---

[8] As the assumption of equal variances across all levels of the repeated-measure variable was violated (perpetrator blaming outgroup $p = .009$), we conducted a robust ANOVA on trimmed means using R Studio Version 1.4.1717 and WRS2 package (Mair & Wilcox, 2020). This analysis revealed similar patterns: a significant main effect for the perpetrator ($p = .045$), no effect of order ($p = .23$), and a significant interaction ($p < .001$).

[9] A robust ANOVA on trimmed means (violation of the assumption of equal variances across all levels of the repeated-measure variable perpetrator blaming outgroup $p < .001$) revealed the same pattern of findings: a significant main effect for the perpetrator ($p = .010$), a significant main effect of order ($p = .017$), and a significant interaction ($p < .001$).

[10] Note that again the assumption of equal variances on the repeated-measure variable was violated (for both ingroup and outgroup classification, $p < .001$), and the robust ANOVA indicated as well a significant interaction between perpetrator times order ($p = .004$) but no main effects (both p values = .169).

[11] The assumption of equal variances for the repeated-measures variable was violated (culture blaming ingroup, $p = .001$), and the robust ANOVA mirrored the observed patterns, such that there was only a main effect of the perpetrator ($p < .001$) but no main effect of order ($p = .187$) nor an interaction effect ($p = .481$).
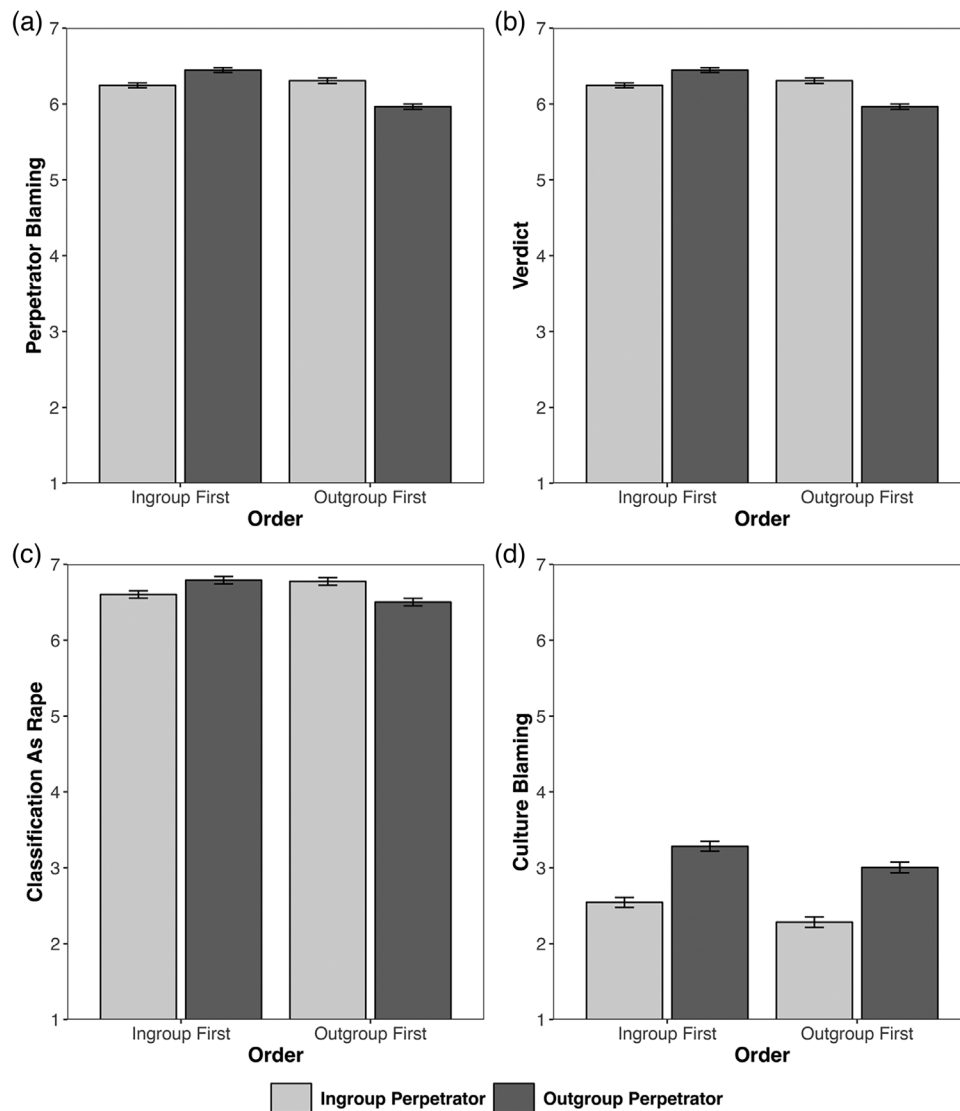
FIGURE 2 *Judgements as a function of perpetrator's group membership and order of presentation. Notes.* a: Perpetrator blaming as a function of the perpetrator's group membership and order of presentation. b: Verdict as a function of perpetrator's group membership and order of presentation. c: Classification as rape as a function of the perpetrator's group membership and order of presentation. d: Culture blaming as a function of the perpetrator's group membership and order of presentation. Higher scores indicate harsher judgements. Error bars indicate the standard error of the mean.

salient (here, by means of confronting participants first with a newsflash about an outgroup perpetrator), we found a clear black sheep effect. With these data, Experiment 2 provides clear evidence for our prediction that the activation of a threatening outgroup stereotype leads to ingroup protection in the form of a black sheep effect. In sharp contrast, participants who saw the ingroup newsflash first expressed harsher judgements towards the outgroup perpetrator. Contrary to the condition in which participants began with the outgroup perpetrator and thus found themselves in an intergroup context, these participants here may have been in an intragroup context, and perhaps even more clearly so than in Experiment 1 (e.g., Haslam et al., 1995). Indeed, when participants who began with the ingroup perpetrator continued with the second newsflash, they encountered an intergroup context. Likely, they then opted to distance themselves from the outgroup perpetra-

tor (as representing a homogenous threatening group) and expressed harsher judgement of him and the outgroup. As far as we know, no work to date has investigated such order effects in the context of the black sheep effect, we followed up with Experiment 3 to examine the robustness of the observed pattern regarding the presentation order.

Experiment 3 further aimed to manipulate participants' perceptions of variability between the ingroup. Prior research revealed that when participants expect less variability between ingroup members (Doosje et al., 1998), then an ingroup deviant comes across even more as an exception and may receive harsher judgements given the threat to the ingroup's image (Lewis & Sherman, 2010; Yzerbyt et al., 2000). To test this conjecture, we examined the impact of a manipulation of ingroup entitativity on ingroup protection at the individual and at the cultural level.

4 | EXPERIMENT 3

Campbell (1958) first introduced the construct of entitativity to refer to the extent to which a group comes across as a real entity (Campbell, 1958; Lickel et al., 2000; Yzerbyt et al., 2000). Entitativity encompasses such aspects as common goals and the degree of interactions and connections between group members (Agadullina & Lovakov, 2017; Campbell, 1958; Yzerbyt et al., 2000). Research suggests that, to the extent that they pose a threat to the entire ingroup, deviant members belonging to highly entitative ingroup (here: fraternity vs. introductory psychology class, in the context of evaluating high- vs. low-quality essays) trigger a black sheep effect (Lewis & Sherman, 2010). Although some work looked at the impact of the entitativity of the ingroup on identification (see Castano et al., 2003, for the relationship between entitativity and identification in the EU context), no research to date manipulated the entitativity of a national group such as Germans to investigate its influence on judgements of highly deviant behaviour. We expected that manipulating ingroup entitativity might foster the effects of ingroup protection that we observed in Experiment 2 both on the individual and the group level.

In Experiment 3, our aims were thus twofold. First, we wanted to test the robustness of the findings of Experiment 2. Specifically, we hoped to replicate a black sheep effect at the individual level and a protective judgement of the ingroup at the cultural level. We further ambitioned to replicate the black sheep effect in the condition when the outgroup newsflash was presented first (intergroup context). Second, we wanted to investigate how people appraise deviance in a context of high (vs. low) ingroup entitativity. Specifically, we predicted a stronger black sheep effect in the high entitativity condition (significant black sheep effect) than in the low entitativity condition (small or no black sheep effect). In addition, we expected less severe condemnation of the ingroup culture in the high entitativity condition than in the low entitativity condition. As such, we tested a 2 (perpetrator: ingroup vs. outgroup) × 2 (order: ingroup first vs. outgroup first) × 2 (entitativity: low vs. high) mixed design with the first factor varying within participants and the two remaining factors between them. As part of Experiment 3, we also adapted the newsflashes regarding guilt certainty, this time decreasing guilt certainty slightly as compared to Experiment 2. We did this because of the overall harsher mean ratings observed in Experiment 2 (e.g., above 5 on a 7-point scale). We reasoned that using the same newsflashes as in Experiment 2 may make it difficult, if not impossible, to observe the effects of our experimental entitativity manipulation (i.e., ceiling effects).

4.1 | Method

4.1.1 | Sample size estimation

We ran a power analysis using the PANGEA (v0.2) web app. For our 2 (perpetrator) × 2 (order) × 2 (entitativity) design, participants were nested in order as well as entitativity and crossed with the perpetrator.

In Experiment 3, we predicted that entitativity should increase the black sheep effect in the high entitativity condition and lead to a decrease or absence of the black sheep effect in the low entitativity condition. We assumed a small effect size, namely $d_z = .20$, for both the black sheep effect (reflected in the main effect of the perpetrator) and the interaction between perpetrator and entitativity. Using this effect size, a total of $N = 700$ participants ($n = 175$ per condition, i.e., order × entitativity) ensures 90% power.

We conducted a sensitivity power analysis (statistical test: ANOVA, repeated measures, within-between interaction) using the final sample size ($N = 703$). For a power of .95 $\alpha = .05$, number of groups = 2 and number of measurements = 2, correlation among repeated measures for perpetrator blaming (culture blaming) = .642 (.571), nonsphericity correction = 1 and the minimum effect size that could be obtained was $f^2 = .057$ (.063).

4.1.2 | Participants and design

A total of $N = 750$ participants from the Prolific Academic with German as their first language participated in the study in two waves. When applying the exclusion criteria, we need to consider both data collection waves to reach our pre-registered sample size of 700 subjects. As pre-registered, in Experiment 3, we excluded participants, if they (a) did not have German as their mother tongue, (b) had parents with a mother tongue from a predominantly Muslim/Arab country, (c) had too short dwell times (≤5 seconds) on the pages with the newsflashes, (d) indicated a wrong answer on the second manipulation check item (i.e., 'what was the conclusion of the results you read at the beginning of the first study?'—targeting the entitativity manipulation) and (e) had extreme values (i.e., too short) in the dwell time on the page where the entitativity manipulation was presented for ensuring that participants read the text about entitativity carefully. Furthermore, prior to data analyses, we excluded participants if (f) the studentized deleted residuals for the dependent measures were > ±4 as recommended by Judd et al. (2011). Based on these exclusion criteria, we excluded 47 participants: $n = 21$ had parents with mother tongues from a predominantly Muslim country, $n = 14$ failed the manipulation check, $n = 5$ (one of them also failed on the manipulation check) spent ≤ 5 seconds on the pages with the vignettes, $n = 2$ who participated in similar previous studies and $n = 6$ based on studentized deleted residuals with absolute values > 4 (Judd et al., 2011). This led to a final sample of $N = 703$[12]. Due to a programming error, participants in the condition of ingroup perpetrator as the first newsflash in the context of cinema did not see the dependent measures regarding verdict items. This is why we only report perpetrator blaming as the dependent measure.

---

[12] Note that the results were generally consistent with the results of analyses without any exclusions (for details, please see the Supplementary Materials).

4.1.3 | Procedure and measures

Overall procedure. We invited participants to take part in two studies. In the first study, participants were randomly assigned to either of the entitativity conditions (high vs. low). Participants read the results of an alleged representative study (see supplementary materials) where, compared to other European citizens, Germans were described as (dis)similar and (not) connected to each other depending on the experimental condition (cf. Crawford et al., 2002, for entitativity manipulations for minimal groups and a group of friends). Following the entitativity manipulation, participants worked through the saying-is-believing task (Bauer & Hannover, 2020; Higgins & Rholes, 1978) and had to come up with three possible reasons for the findings they read about. This aimed to strengthen the entitativity manipulation.

Participants then went on to the alleged second study. Here the procedure was the same as in Experiment 2. After working through the first newsflash and its blame judgements, participants worked through the second newsflash and the blame items and ended the study by indicating some demographics before being fully debriefed. We included two attention checks throughout the study (e.g., 'Please select the 4") and warned participants about their presence at the beginning of the study. We further added one manipulation check item examining whether participants correctly remembered the content of the first article that was crucial for the entitativity manipulation: 'What was the conclusion of the results you read in the beginning of this study?' (1 = *Germans were described as similar to each other compared to other European countries*, 2 = *Germans were described as dissimilar to each other compared to other European countries*). This item came after the end of the alleged Study 2 to avoid influencing participants directly after the entitativity study. In addition, we asked one explorative item as follows: 'How similar and connected do you believe are Germans?' (a scale ranging from 1 = *not at all similar* to 7 = *very similar*).

Newsflashes. We used the newsflashes from Experiment 2 and, as mentioned above, adapted the last sentence of each newsflash for leaving room for guilt being slightly more ambiguous. For instance, in the dating context, we adapted the concluding sentence as follows: 'According to the police report, he initially denied the crime. But after several days of lengthy interrogations, he confessed to the rape'. In the cinema context, the concluding sentence read as follows: 'Andreas S. (Ahmed S.) finally confessed to the crime after the interrogations by the police even though he was convinced by the actions being consensual until the end'.

Judgements. The dependent measures were the same as in Experiment 2.

4.2 | Results

We analysed our data with 2 (perpetrator: ingroup vs. outgroup) × 2 (order: ingroup first vs. outgroup first) × 2 (entitativity: high vs. low) mixed-model ANOVAs.

4.2.1 | Individual Blame

Perpetrator blaming. The first analysis with perpetrator blaming[13] as a dependent variable revealed a significant main effect of the perpetrator, $F(1,699) = 57.45$, $p < .001$, $\eta_p^2 = .08$ and 95% CI [0.04–0.12], indicating that participants blamed the ingroup perpetrator more strongly ($M = 6.15, SD = 0.78$) than his outgroup counterpart ($M = 5.95$, $SD = 0.92$). We further observed a significant main effect of order $F(1,699) = 10.62$, $p = .001$, $\eta_p^2 = .01$ and 95% CI [0.00–0.04], such that the overall judgements of the perpetrators were harsher when the ingroup came first ($M = 6.14, SD = 0.80$) compared to when the outgroup came first ($M = 5.96, SD = 0.90$). In addition, we observed a significant interaction of perpetrator × order, $F(1,699) = 99.09$, $p < .001$, $\eta_p^2 = .12$ and 95% CI [0.08–0.17]. To probe the interaction, we ran simple effects for the perpetrator, separately for each order. When the outgroup newsflash came first, the ingroup perpetrator was blamed more harshly ($M = 6.18, SD = 0.76$) than the outgroup perpetrator ($M = 5.73, SD = 0.97$), $t(699) = -12.48$, $p < .001$, $\eta_p^2 = .18$ and 95% CI [0.13–0.23]. When the ingroup newsflash came first, there was no significant difference between the blame attributions. However, the outgroup perpetrator tended to be blamed more harshly ($M = 6.17$, $SD = 0.81$) compared to his ingroup counterpart ($M = 6.12, SD = 0.81$), $t(699) = 1.67$, $p = .10$, $\eta_p^2 = .004$ and 95% CI [0.00–0.02]. We did not observe any significant main or interaction effect regarding entitativity, see Figure 3, panel a. Results concerning the verdict measure can be found as part of the S1. It is noteworthy, however, that the manipulation check indicated that the entitativity manipulation was successful as it affected the degree of similarity and connectedness perceived by participants, $F(1, 701) = 87.39$, $p < .001$ and $\eta_p^2 = .11$. Participants in the entitativity high condition ($M = 4.57, SD = 1.22$) indicated higher degrees of perceived similarity and connectedness between Germans than those in the entitativity low condition ($M = 3.73, SD = 1.17$).

Classification as rape. The analysis of the classification of the perpetrator's behaviour as rape[14] revealed a significant effect of the perpetrator, $F(1,699) = 49.22$, $p < .001$, $\eta_p^2 = .07$ and 95% CI [0.03–0.10], in that the ingroup perpetrator's behaviour was classified more likely as rape ($M = 6.54, SD = 0.89$) than the outgroup perpetrator's behaviour ($M = 6.21, SD = 1.15$). We further observed a significant effect of order, $F(1,699) = 5.98$, $p = .015$, $\eta_p^2 = .008$ and 95% CI [0.00–0.03]. Again, the overall judgements for the classification as rape were harsher when the ingroup newsflash came first ($M = 6.45, SD = 0.99$) than when the outgroup newsflash came first ($M = 6.31, SD = 1.09$). In addition, we observed a significant interaction between perpetrator and order, $F(1,699) = 26.45$, $p < .001$, $\eta_p^2 = .04$ and 95% CI [0.01–0.07]. Follow-up tests for each order separately revealed that, when the outgroup

---

[13] Because the assumption of homogeneity of variances was violated for the repeated measures variable (perpetrator blaming outgroup $p = .001$) we also ran a robust ANOVA on trimmed means as in Exp. 2. This analysis revealed the same pattern of results with both significant main effects (perpetrator: $p < .001$, order: $p = .004$) and the interaction effect ($p < .001$).

[14] Because the assumption of homogeneity of variances was violated for the repeated measures variable (classification as rape ingroup $p = .001$, outgroup $p < .001$) we also conducted a robust ANOVA on trimmed means which was in line with the observed significant results (perpetrator: $p < .001$, order: $p = .009$) interaction effect ($p < .001$).
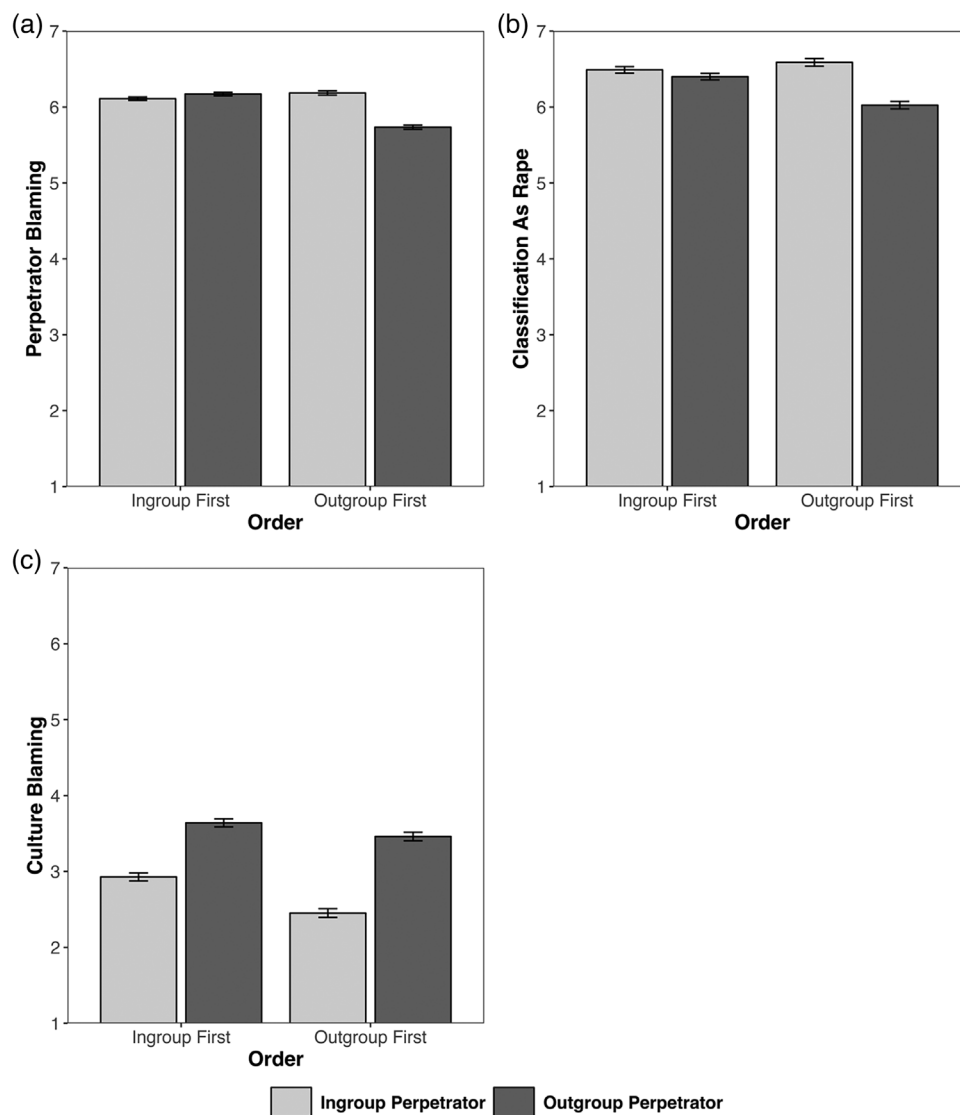
FIGURE 3 *Judgements as a function of perpetrator's group membership and order of presentation. Notes.* a: Perpetrator blaming as a function of the perpetrator's group membership and order of presentation. b: Classification as rape as a function of the perpetrator's group membership and order of presentation. c: Culture blaming as a function of the perpetrator's group membership and order of presentation. Higher scores indicate harsher judgements. Error bars indicate the standard error of the mean.

newsflash came first, the ingroup perpetrator's behaviour was classified more as rape ($M = 6.59, SD = 0.79$) than the outgroup perpetrator's behaviour ($M = 6.03, SD = 1.25$), $t(699) = -8.66, p < .001, \eta_p^2 = .10$ and 95% CI [0.06–0.14]. When the ingroup newsflash came first, there was no significant difference between the classification as rape ($M = 6.49, SD = 0.98$ and $M = 6.40, SD = 0.99$, for ingroup and outgroup perpetrator, respectively), $t(699) = -1.32, p = .19$ (see Figure 3, panel b). We did not find any significant main or interaction effect regarding entitativity.

Culture blaming. We again replicated the predicted effect of the perpetrator on culture blaming, $F(1,699) = 242.05, p < .001, \eta_p^2 = .26$ and 95% CI [0.21–0.31], such that participants judged the ingroup culture more leniently ($M = 2.68, SD = 1.47$) than the outgroup culture ($M = 3.55, SD = 1.68$). We also observed an order effect, $F(1,699) = 8.98, p = .003, \eta_p^2 = .01$ and 95% CI [0.00–0.03], such that the overall judgements were harsher when the ingroup newsflash came

first ($M = 3.28, SD = 1.67$ and $M = 2.96, SD = 1.59$, for ingroup and outgroup newsflash, respectively). In addition, we found a significant interaction[15] between perpetrator × order, $F(1,699) = 6.83, p = .009, \eta_p^2 = .01$ and 95% CI [0.00–0.03] (see Figure 3, panel c). Simple effects of perpetrator for each order separately indicated that, when the outgroup newsflash came first, the ingroup culture was judged more leniently ($M = 2.45, SD = 1.33$) than the outgroup culture ($M = 3.46, SD = 1.68$), $t(699) = -12.94, p < .001, \eta_p^2 = .19$ and 95% CI [0.14–0.24]. Similarly, when the ingroup newsflash came first, the ingroup culture was again judged more leniently ($M = 2.93, SD = 1.57$) than the outgroup culture ($M = 3.64, SD = 1.69$) ($M = 2.93, SD = 1.57$), $t(699) = 9.09$,

[15] Because the assumption of homogeneity of variances was violated for the repeated measures variable (culture blaming ingroup $p = .006$) we also conducted a robust ANOVA on trimmed means. This indicated both significant main effects (perpetrator: $p < .001$, order: $p = .003$) but a marginal interaction effect ($p = .106$).

$p < .001$, $\eta_p^2 = .11$ and 95% CI [0.07–0.15]. We did not observe any significant main or interaction effect regarding entitativity.

4.3 | Discussion

Experiment 3 confirmed the predicted patterns in terms of the black sheep effect for perpetrator blaming and classification as rape. Participants indicated harsher blame towards the ingroup than the outgroup perpetrator. They also considered the perpetrator's behaviour more as rape when he belonged to the ingroup than to the outgroup. More importantly, we replicated the black sheep effect when the outgroup newsflash was presented first. That is, participants expressed harsher judgements towards the ingroup perpetrator when they had first read the outgroup newsflash. This pattern was present also in the classification of the perpetrator's behaviour as rape as participants perceived the ingroup perpetrator's behaviour more as rape when they first worked through the outgroup newsflash. This is in line with our view that the activation of a threatening outgroup before the judgement of an ingroup member elicits harsher judgements of the ingroup deviant. Regarding the blame attributed towards the culture of the perpetrators, we replicated the exoneration of the ingroup culture. We further observed a higher difference between the culture blame judgements when the outgroup newsflash was presented first. Again, this confirms that making an intergroup context salient reinforces group members' reaction to protect the ingroup (Marques & Yzerbyt, 1988).

As far as entitativity was concerned and despite the apparent success of our manipulation, we did not observe any main nor interaction effects involving this factor. One possible explanation may be that the seriousness of the deviant behaviour selected in the present context prevents the emergence of any visible impact of this factor. Furthermore, it has been shown that intimacy groups receive the highest entitativity ratings followed by task groups, social groups and loose associations (Denson et al., 2006; Lickel et al., 2000). As such, it may be challenging to manipulate the entitativity of a social group such as Germans. Previous work used minimal groups, friends, experimental confederates with similar field-hockey sweaters (e.g., Crawford et al., 2002; Pereira & van Prooijen, 2018) or groups where the cohesiveness of the group was perceived as high without further manipulation, that is, fraternities (Lewis & Sherman, 2010). It thus remains to be seen how entitativity or other factors affecting the subjective homogeneity of the ingroup (or of the outgroup) may play a role in the emergence of our predicted pattern on individual-level and group-level judgements.

5 | GENERAL DISCUSSION

In the present series of experiments, we investigated ingroup protection at both the individual and cultural levels. We built on recent work showing harsher judgements of individual ingroup perpetrators (i.e., black sheep effect) on the one hand and more lenient judgement of the ingroup as a whole on the other (Khosrowtaj et al., 2024c). In Experiment 1, we tested whether the mere salience of a threatening outgroup may elicit harsher judgements of an ingroup than an outgroup perpetrator using a purely intragroup design. Experiment 2 built on Experiment 1 and highlighted the importance of an intergroup context using a within-participant manipulation of the perpetrator's group membership. In Experiment 3, we tested the robustness of the patterns observed in Experiment 2 and further tested the effect of high (vs. low) ingroup entitativity on judging an ingroup perpetrator and his culture.

5.1 | Lessons Learned

The present work found support for the black sheep effect at the individual perpetrator level, that is, participants judged the perpetrator more harshly when he was an ingroup than an outgroup member (Marques et al., 1988; Marques & Yzerbyt, 1988). Our distinct contribution to the black sheep effect literature concerns the examination of people's tendency to protect the ingroup as a whole (Marques, 1990; Marques & Yzerbyt, 1988). Participants blamed the ingroup culture (including its norms and values) less severely for the deviant behaviour than they blamed the outgroup culture. As far as we know, the present contribution is the first to show the protection of the ingroup both on individual and cultural levels. That is, in two of three experiments, we observed the harsher condemnation of one ingroup member. While all three experiments found that the ingroup was protected culturally. To be sure, these results may also correspond to a derogation of the outgroup. Both judgements on the individual and cultural levels are comparative, that is, one may take the perspective that one outgroup perpetrator is not judged more harshly than his ingroup counterpart, while on the cultural level, participants consider the culture of the outgroup perpetrator to be more responsible for the deviant behaviour.

The order effects observed in Experiment 2 are fully consistent with the idea underlying our outgroup salience manipulation of Experiment 1. When participants face an ingroup perpetrator *after* being reminded of the threatening outgroup (which is activated by the deviant behaviour of the outgroup perpetrator in the first newsflash), they distanced themselves more extremely from the ingroup deviant. This finding proved to be robust as we replicated this pattern in Experiment 3. This suggests that the presence of an intergroup context may well constitute a necessary condition for the black sheep effect to occur.

This difference between the ascription of blame at the individual and the cultural level hints to a possible proximity to the concept of shifting standards (e.g., Biernat & Manis, 1994). Different judgement standards may have been used for in- and outgroup deviants based on group stereotypes (Biernat et al., 1991). Even though the present work did not focus on the use of different dependent variables (i.e., subjective vs. objective items), the black sheep effect may be seen as resulting from the more demanding standard for the ingroup perpetrator and more favourable judgement (leniency) towards the outgroup perpetrator based on the stereotype that 'Muslims/Afghans are violent' (Biernat et al., 1991; Linville & Jones, 1980). Indeed, shifting standards have been observed to be motivated by ingroup protection. In one illustrative study (Miron et al., 2010), participants were less motivated to protect the ingroup and experienced higher

collective guilt with respect to America's history of slavery after they had affirmed their group's value. Importantly, all dependent measures used by Miron and colleagues (2010) targeted only the ingroup. As such, the present contribution extends the examination of different standards as we compared judgements attributed to both in- and outgroup perpetrators on individual and cultural levels.

5.2 | Limitations and Future Directions

Regarding the failure of our priming manipulation, we suggest that the operationalization was possibly too broad (Ledgerwood & Chaiken, 2007) and thus too weak to activate the idea of a threatening outgroup. Narrower primes may elicit the intergroup context and lead to harsher black sheep effects, which have been discussed alongside the order effects of Experiments 2–3. Along similar lines, our entitativity manipulation of Experiment 3 did not affect participants' judgements of the perpetrators and their cultures. At the same time, participants in the high entitativity condition did perceive Germans as more similar and connected to each other than participants in the low entitativity condition. This suggests that the absence of the predicted pattern may also be due to a ceiling effect. As much as this, the idea behind underlying both the outgroup salience and entitativity manipulations remains worth pursuing, and future research may address these aspects more explicitly. Future work may also benefit from considering the perceptions of outgroup entitativity. If participants indeed perceive the outgroup as entitative and threatening (Agadullina & Lovakov, 2017; Sacchi et al., 2009; Vasquez et al., 2015; Yzerbyt et al., 2000), this may motivate them more to distance themselves from an ingroup deviant and to judge their own culture more leniently.

One may further ask whether the differences in culture blaming arise as the outcome of ingroup protection or outgroup derogation. One strategy to approach this issue may rest on a comparison between the influence of ingroup identification on the one hand and of outgroup prejudice on the other. To the extent that these two aspects can be separated, future work would benefit from distinguishing between the two processes. Interestingly, recent work hints at the existence of a link between national narcissism and hostile intentions against refugees and immigrants. Specifically, national narcissism but not national identification predicted hostile attitudes as well as collective action against refugees and immigrants (Górska et al., 2022). This is an intriguing finding and further investigation of this pattern would indeed be a promising avenue for future research. The issue concerns the underlying mechanisms of the ingroup protection at the individual and group levels.

The present work may also inspire future efforts to investigate whether ingroup protection would emerge on both individual and cultural levels in other cultural and national contexts. We assume similar effects in the context of relevant stigmatized groups when the violated norm is of relevance for the ingroup (Brewer, 2001; Marques, 1990), but this remains an open question for future empirical work.

Finally, and leaving the question of the way people appraise negative acts, it would be fruitful to examine whether positive acts (such as success) from ingroup and outgroup members are attributed to their culture. Future research may follow this promising research idea for raising awareness for differential treatments and portrayals of foreigners, residents with immigration backgrounds and refugees.

5.3 | Conclusions

The present work offers several interesting insights into the existing literature. First, we show that in spite of the impression that may stem from existing media debates where negative attitudes and hostility towards Muslims are prevalent (Bauer & Hannover, 2020; Stürmer et al., 2019; Zick et al., 2011), German participants judge more harshly a German perpetrator than his Afghan counterpart. This black sheep effect may seem paradoxical, but it essentially preserves the positivity of the ingroup as a whole (Marques & Yzerbyt, 1988). Second, our data show that it matters whether participants find themselves in an intragroup or an intergroup context. Making a threatening outgroup salient facilitates the emergence of the black sheep effect as a way to express that such a deviant behaviour is not typical for the ingroup (Marques & Yzerbyt, 1988; Yzerbyt et al., 2000). Third, and most importantly, going beyond the classic black sheep effect, the ingroup is also protected on the cultural level. The comparative leniency towards the ingroup culture may be a form of outgroup discrimination on the cultural level, which allows incriminating all outgroup members. When the perpetrator is an ingroup member, it is only the individual harm-doer who is judged more harshly (i.e., black sheep effect), thereby deflecting the responsibility of the rest of the ingroup. When facing an outgroup perpetrator, it is not him but his culture that ingroup members derogate. These data suggest that the outgroup deviant is not so much treated as an individual but rather as an interchangeable instance of an otherwise homogenous and indeed negative outgroup. Raising the level of awareness regarding this judgemental strategy may be crucial if one wishes to decrease hostility against Muslims and reduce anti-Muslim prejudice (Bruneau et al., 2020; Gallardo et al., 2021). Considering the potential consequences on intergroup relations, investigating this aspect is undoubtedly a promising avenue for future research.

AUTHOR CONTRIBUTION
All authors contributed substantially to this work. More precisely, Sarah Teige-Mocigemba contributed to the conception and design, supported data acquisition, supported the analysis and interpretation of data and was involved in revising the manuscript critically for its intellectual content. Vincent Yzerbyt contributed to the conception and design, supported the analysis and interpretation of data and was involved in revising the manuscript critically for its intellectual content. Zahra Khosrowtaj was involved in conception, design, data acquisition, data analysis and interpretation of data and drafting the manuscript.

CONFLICT OF INTEREST STATEMENT

No potential conflict of interest was reported by the authors.

ETHICS STATEMENT

The ethical principles as formulated in the WMA Declaration of Helsinki guided our research project. If research objectives do not involve issues regulated by law (e.g., the German Medicine Act [Arzneimittelgesetz, AMG], the Medical Devices Act [Medizinproduktegesetz, MGP], the Stem Cell Research Act [Stammzellenforschungsgesetz, StFG] or the Medical Association's Professional Code of Conduct [Berufsordnung der Ärzte]), then no ethics approval is required for social science research in Germany. Our studies have no such objectives, and therefore, no IRB approval or waiver of permission was sought for these studies.

TRANSPARENCY STATEMENT

We pre-registered all the projects and the data as well as the analysis syntax for all three experiments are available at OSF (DOI 10.17605/OSF.IO/3MGJ5): https://osf.io/3mgj5/

ORCID

Zahra Khosrowtaj https://orcid.org/0000-0001-5327-9656
Sarah Teige-Mocigemba https://orcid.org/0000-0002-7516-1293
Vincent Yzerbyt https://orcid.org/0000-0003-1185-4733

REFERENCES

Abrams, D., Marques, J., Bown, N., & Dougill, M. (2002). Anti-norm and pro-norm deviance in the bank and on the campus: Two experiments on subjective group dynamics. *Group Processes & Intergroup Relations*, 5(2), 163–182. https://doi.org/10.1177/1368430202005002922

Abrams, D., Marques, J. M., Bown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology*, 78(5), 5. https://doi.org/10.1037/0022-3514.78.5.906

Agadullina, E., & Lovakov, A. (2017). Understanding entitativity: Are there real differences between approaches? *Psychology. Journal of Higher School of Economics*, 14, 536–554. https://doi.org/10.17323/1813-8918-2017-3-536-554

Bauer, C. A., & Hannover, B. (2020). Changing "us" and hostility towards "them"—Implicit theories of national identity determine prejudice and participation rates in an anti-immigrant petition. *European Journal of Social Psychology*, 50(4), 810–826. https://doi.org/10.1002/ejsp.2666

Becker, L. (2023, January 11). *Merz nennt Migrantenkinder 'kleine Paschas'*. WDR. https://www1.wdr.de/nachrichten/merz-nennt-migranten-kinder-kleine-paschas-100.html

Bettencourt, B. A., Dill, K. E., Greathouse, S. A., Charlton, K., & Mulholland, A. (1997). Evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology*, 33(3), Article 3. https://doi.org/10.1006/jesp.1996.1323

Bieneck, S., & Krahé, B. (2011). Blaming the victim and exonerating the perpetrator in cases of rape and robbery: Is there a double standard? *Journal of Interpersonal Violence*, 26(9), Article 9. https://doi.org/10.1177/0886260510372945

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66, 5–20. https://doi.org/10.1037/0022-3514.66.1.5

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60, 485–499. https://doi.org/10.1037/0022-3514.60.4.485

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17(5), 475–482. https://doi.org/10.1177/0146167291175001

Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In *Social identity, intergroup conflict, and conflict reduction* (pp. 17–41). Oxford University Press.

Bridges, J. S. (1991). Perceptions of date and stranger rape: A difference in sex role expectations and rape-supportive beliefs. *Sex Roles*, 24(5), Article 5. https://doi.org/10.1007/BF00288303

Bruneau, E. G., Kteily, N. S., & Urbiola, A. (2020). A collective blame hypocrisy intervention enduringly reduces hostility towards Muslims. *Nature Human Behaviour*, 4(1), Article 1. https://doi.org/10.1038/s41562-019-0747-7

Campbell, D. (1958). Common fate, similarity, and other indexes of the status of aggregates of persons as social entities. *Behavioral Science*, 3(1), Article 1. https://doi.org/10.1002/bs.3830030103

Carnaghi, A., & Yzerbyt, V. Y. (2007). Subtyping and social consensus: The role of the audience in the maintenance of stereotypic beliefs. *European Journal of Social Psychology*, 37(5), Article 5. https://doi.org/10.1002/ejsp.402

Castano, E., Yzerbyt, V., & Bourguignon, D. (2003). We are one and I like it: The impact of ingroup entitativity on ingroup identification. *European Journal of Social Psychology*, 33(6), Article 6. https://doi.org/10.1002/ejsp.175

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, 83(5), Article 5. https://doi.org/10.1037//0022-3514.83.5.1076

Czymara, C. S., & Schmidt-Catran, A. W. (2017). Refugees unwelcome? Changes in the public acceptance of immigrants and refugees in Germany in the course of Europe's 'Immigration Crisis'. *European Sociological Review*, 33(6), Article 6. https://doi.org/10.1093/esr/jcx071

Denson, T. F., Lickel, B., Curtis, M., Stenstrom, D. M., & Ames, D. R. (2006). The roles of entitativity and essentiality in judgments of collective responsibility. *Group Processes & Intergroup Relations*, 9(1), Article 1. https://doi.org/10.1177/1368430206059857

Di Stasio, V., Lancee, B., Veit, S., & Yemane, R. (2019). Muslim by default or religious discrimination? Results from a cross-national field experiment on hiring discrimination. *Journal of Ethnic and Migration Studies*, 47(6), 1305–1326. https://doi.org/10.1080/1369183X.2019.1622826

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, 75(4), Article 4. https://doi.org/10.1037/0022-3514.75.4.872

Duncan, B. (1976). Differential social-perception and attribution of intergroup violence—Testing lower limits of stereotyping of Blacks. *Journal*

of Personality and Social Psychology, *34*(4), Article 4. https://doi.org/10.1037/0022-3514.34.4.590

Dunwoody, P. T., & McFarland, S. G. (2018). Support for anti-Muslim policies: The role of political traits and threat perception. *Political Psychology*, *39*(1), 89–106. https://doi.org/10.1111/pops.12405

Ferrin, M., Mancosu, M., & Cappiali, T. M. (2020). Terrorist attacks and Europeans' attitudes towards immigrants: An experimental approach. *European Journal of Political Research*, *59*(3), Article 3. https://doi.org/10.1111/1475-6765.12362

Fitzgerald, J., Curtis, K. A., & Corliss, C. L. (2012). Anxious publics: Worries about crime and immigration. *Comparative Political Studies*, *45*(4), 477–506. https://doi.org/10.1177/0010414011421768

Gallardo, R. A., Hameiri, B., Moore-Berg, S. L., & Bruneau, E. (2021). The collective praise intervention: A brief intervention highlighting prosocial behavior reduces hostility towards Muslims. *Group Processes & Intergroup Relations*, *25*(7), 1696–1717. https://doi.org/10.1177/13684302211021665

George, W. H., & Martínez, L. J. (2002). Victim blaming in rape: Effects of victim and perpetrator race, type of rape, and participant racism. *Psychology of Women Quarterly*, *26*(2), Article 2. https://doi.org/10.1111/1471-6402.00049

Górska, P., Stefaniak, A., Marchlewska, M., Matera, J., Kocyba, P., Łukianow, M., Malinowska, K., & Lipowska, K. (2022). Refugees unwelcome: Narcissistic and secure national commitment differentially predict collective action against immigrants and refugees. *International Journal of Intercultural Relations*, *86*, 258–271. https://doi.org/10.1016/j.ijintrel.2021.11.009

Gümüsay, K. (2022). *Speaking and being: How language binds and frees us* (I. Gesche, Trans.). Profile Books.

Gusciute, E., Muhlau, P., & Layte, R. (2021). All welcome here? Attitudes towards Muslim migrants in Europe. *International Migration*, *59*(5), Article 5. https://doi.org/10.1111/imig.12810

Haslam, S. A., & Oakes, P. J. (1995). How context-independent is the outgroup homogeneity effect? A response to Bartsch and Judd. *European Journal of Social Psychology*, *25*(4), 469–475. https://doi.org/10.1002/ejsp.2420250410

Haslam, S. A., Oakes, P. J., Turner, J. C., & McGarty, C. (1995). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology*, *34*(2), Article 2. https://doi.org/10.1111/j.2044-8309.1995.tb01054.x

Hestermann, T. (2019). *Berichterstattung über Gewaltkriminalität: Wie häufig nennen Medien die Herkunft von Tatverdächtigen?* [Berlin: Mediendienst Integration]. Mediendienst Integration.

Hewstone, M. (1990). The 'ultimate attribution error'? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology*, *20*, 311–335. https://doi.org/10.1002/ejsp.2420200404

Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*(4), 363–378. https://doi.org/10.1016/0022-1031(78)90032-X

Hirtenlehner, H. (2019). Gefährlich sind immer die Anderen!: Migrationspanik, Abstiegsängste und Unordnungswahrnehmungen als Quelle der Furcht vor importierter KriminalitätAnxieties about immigration of refugees, fears of falling and perceptions of disorder as sources of worries about imported crime. *Monatsschrift für Kriminologie und Strafrechtsreform*, *102*(4), 262–281. https://doi.org/10.1515/mks-2019-2026

Islam, M. R., & Hewstone, M. (1993). Intergroup attributions and affective consequences in majority and minority groups. *Journal of Personality and Social Psychology*, *64*, 936–950. https://doi.org/10.1037/0022-3514.64.6.936

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2011). *Data analysis: A model comparison approach* (2nd ed.). Routledge.

Kauff, M. (2024). Muslim = Terrorist? Attribution of violent crimes to terrorism or mental health problems depend on perpetrators' religious background. *The Journal of Social Psychology*, *164*(4), 447–455. https://doi.org/10.1080/00224545.2022.2095968

Kearns, E. M., Betus, A. E., & Lemieux, A. F. (2019). Why do some terrorist attacks receive more media attention than others? *Justice Quarterly*, *36*(6), Article 6. https://doi.org/10.1080/07418825.2018.1524507

Khosrowtaj, Z., Biermann, V., & Teige-Mocigemba, S. (2024a). "He must be a foreigner" - On biased information search towards in- and outgroup wrongdoers. *Manuscript in Preparation*, *00*, 00–00.

Khosrowtaj, Z., Süssenbach, P., & Teige-Mocigemba, S. (2024b). The black sheep effect as a shift of blame? Exonerating individual outgroup perpetrators at the expense of the victim and the outgroup as a whole. *Manuscript Submitted for Publication*, *00*, 00–00.

Khosrowtaj, Z., Yzerbyt, V., Süssenbach, P., & Teige-Mocigemba, S. (2024c). On the effect of perpetrator's description on inter- and intragroup guilt attributions. *Manuscript in Preparation*, *00*, 00–00.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), Article 5. https://doi.org/10.1037/pspp0000048

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*(4), Article 4. https://doi.org/10.1037/0022-3514.68.4.565

Ledgerwood, A., & Chaiken, S. (2007). Priming us and them: Automatic assimilation and contrast in group attitudes. *Journal of Personality and Social Psychology*, *93*(6), Article 6. https://doi.org/10.1037/0022-3514.93.6.940

Lewis, A. C., & Sherman, S. J. (2010). Perceived entitativity and the black-sheep effect: When will we denigrate negative ingroup members? *The Journal of Social Psychology*, *150*, 211–225. https://doi.org/10.1080/00224540903366388

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*(2), Article 2. https://doi.org/10.1037/0022-3514.78.2.223

Liebe, U., Meyerhoff, J., Kroesen, M., Chorus, C., & Glenk, K. (2018). From welcome culture to welcome limits? Uncovering preference changes over time for sheltering refugees in Germany. *PLOS ONE*, *13*(8), Article 8. https://doi.org/10.1371/journal.pone.0199923

Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of out-group members. *Journal of Personality and Social Psychology*, *38*, 689–703. https://doi.org/10.1037/0022-3514.38.5.689

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Mange, J., Chun, W. Y., Sharvit, K., & Belanger, J. J. (2012). Thinking about Arabs and Muslims makes Americans shoot faster: Effects of category accessibility on aggressive responses in a shooter paradigm. *European Journal of Social Psychology*, *42*(5), 552–556. https://doi.org/10.1002/ejsp.1883

Marques, J. M. (1990). The black sheep effect: Outgroup homogeneity as a social comparison process. In *Social identity theory: Constructive and critical advances* (pp. 131–151). Harvester Wheatsheaf.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgmental extremity towards ingroup members in inter- and intra-group situations. *European Journal of Social Psychology*, *18*(3), Article 3. https://doi.org/10.1002/ejsp.2420180308

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

Miron, A. M., Branscombe, N. R., & Biernat, M. (2010). Motivated shifting of justice standards. *Personality and Social Psychology Bulletin*, *36*(6), 768–779. https://doi.org/10.1177/0146167210370031

Moll, M. (2023, January 5). *Berliner CDU fragt nach Vornamen der Tatverdächtigen—Linke erhebt Vorwürfe*. FOCUS online. https://www.focus.de/politik/deutschland/nach-berliner-silvesterkrawallen-stellen-christdemokraten-im-berliner-senat-das-deutschsein-der-deutschen-tatverdaechtigen-in-frage_id_182220787.html

Otten, S., & Gordijn, E. H. (2014). Was it one of us? How people cope with misconduct by fellow ingroup members. *Social and Personality Psychology Compass*, *8*(4), Article 4. https://doi.org/10.1111/spc3.12098

Parks-Stamm, E. J. (2013). Exclusion as self-protection: The function of sub-types for ingroup members. *Personality and Social Psychology Bulletin*, *39*(5), Article 5. https://doi.org/10.1177/0146167213479978

Pereira, A., & van Prooijen, J.-W. (2018). Why we sometimes punish the inno-cent: The role of group entitativity in collective punishment. *PloS One*, *13*(5), Article 5. https://doi.org/10.1371/journal.pone.0196852

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, *5*(4), Article 4. https://doi.org/10.1177/014616727900500407

Sacchi, S., Castano, E., & Brauer, M. (2009). Perceiving one's nation: Entitativity, agency and security in the international arena. *International Journal of Psychology*, *44*(5), Article 5. https://doi.org/10.1080/00207590802236233

Seelig, L. (2023, January 13). Karim Fereidooni: 'Eltern haben Angst um den Bildungserfolg ihres kleinen Maximilian'. *Zeit Magazin*. https://www.zeit.de/zeit-magazin/familie/2023-01/karim-fereidooni-silvesterkrawalle-rassismus-schulen-forschung

Shooman, Y. (2012). *Das Zusammenspiel von Kultur, Religion, Ethnizität und Geschlecht im antimuslimischen Rassismus | APuZ*. bpb.de. http://www.bpb.de/apuz/130422/das-zusammenspiel-von-kultur-religion-ethnizitaet-und-geschlecht-im-antimuslimischen-rassismus

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Mus-lim immigration, critical events, and the seeds of majority members' support for radical responses: An interactionist perspective. *Personality and Social Psychology Bulletin*, *45*(1), Article 1. https://doi.org/10.1177/0146167218780989

Süssenbach, P., Bohner, G., & Eyssel, F. (2012). Schematic influences of rape myth acceptance on visual information processing: An eye-tracking approach. *Journal of Experimental Social Psychology*, *48*(3), Article 3. https://doi.org/10.1016/j.jesp.2011.12.005

Süssenbach, P., Eyssel, F., Rees, J., & Bohner, G. (2017). Looking for blame: Rape myth acceptance and attention to victim and perpetrator. *Journal of Interpersonal Violence*, *32*(15), Article 15. https://doi.org/10.1177/0886260515591975

Tagesschau. (2023a, January 3). *Übergriffe in der Silvesternacht: 103 Personen wieder auf freiem Fuß*. tagesschau.de. https://www.tagesschau.de/inland/innenpolitik/silvester-krawalle-festnahmen-101.html

Tagesschau. (2023b, January 3). *Übergriffe in der Silvesternacht: Diskussion über 'gescheiterte Integration'*. tagesschau.de. https://www.tagesschau.de/inland/silvester-gewalt-gegen-polizisten-101.html

Taylor, D. M., & Jaggi, V. (1974). Ethnocentrism and causal attribution in a South Indian context. *Journal of Cross-Cultural Psychology*, *5*(2), Article 2. https://doi.org/10.1177/002202217400500202

Unkelbach, C., Forgas, J. P., & Denson, T. F. (2008). The turban effect: The influence of Muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology*, *44*(5), Article 5. https://doi.org/10.1016/j.jesp.2008.04.003

Unkelbach, C., Schneider, H., Gode, K., & Senft, M. (2010). A Turban effect, too: Selection biases against women wearing Muslim headscarves. *Social Psychological and Personality Science*, *1*(4), Article 4. https://doi.org/10.1177/1948550610378381

van Prooijen, J.-W. (2006). Retributive reactions to suspected offenders: The importance of social categorizations and guilt probability. *Personality & Social Psychology Bulletin*, *32*(6), Article 6. https://doi.org/10.1177/0146167205284964

Vasquez, E. A., Wenborne, L., Peers, M., Alleyne, E., & Ellis, K. (2015). Any of them will do: In-group identification, out-group entitativity, and gang membership as predictors of group-based retribution. *Aggressive Behavior*, *41*(3), Article 3. https://doi.org/10.1002/ab.21581

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual vio-lence: 'Intersectional stereotyping' in mass media representations of male Muslim migrants in Germany. *Culture and Religion*, *20*(3), Article 3. https://doi.org/10.1080/14755610.2019.1658609

Wigger, I., Yendell, A., & Herbert, D. (2022). The end of 'Welcome Culture'? How the Cologne assaults reframed Germany's immigration discourse. *European Journal of Communication*, *37*(1), Article 1. https://doi.org/10.1177/02673231211012173

Windisch, M., AFP, & Schmalz, A. (2023, January 3). *Silvester-Randale: Berliner Polizei nennt Nationalitäten der Festgenommenen*. Berliner Zeitung. https://www.berliner-zeitung.de/news/silvester-in-berlin-ueber-100-randalierer-wieder-auf-freiem-fuss-li.303354

Yuki, M., & Yokota, K. (2009). The primal warrior: Outgroup threat prim-ing enhances intergroup discrimination in men but not women. *Journal of Experimental Social Psychology*, *45*(1), 271–274. https://doi.org/10.1016/j.jesp.2008.08.018

Yzerbyt, V., Castano, E., Leyens, J.-P., & Paladino, M.-P. (2000). The primacy of the ingroup: The interplay of entitativity and identification. *European Review of Social Psychology*, *11*(1), Article 1. https://doi.org/10.1080/14792772043000059

Zick, A., Küpper, B., & Hövermann, A. (2011). *Intolerance, prejudice and discrimination: A European report*. Friedrich-Ebert-Stiftung.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Support-ing Information section at the end of this article.