On the evaluation of intergroup deviance on individual and group level:

“When it is (n)one of us, but all of them.”

Dissertation

„kumulativ“

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Psychologie der Philipps-Universität Marburg

Vorgelegt von

M.Sc., Zahra Khosrowtaj

Aus Teheran, Iran

Marburg, 2023

Vom Fachbereich Psychologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 09.07.2023

Erstgutachterin: Prof. Dr. Sarah Teige-Mocigemba

Zweitgutachter: Prof. Dr. Vincent Yzerbyt

Tag der Disputation: 13.07.2023

Index

List of Tables

INTRODUCTION OF THE THESIS:

MANUSCRIPT 1:

MANUSCRIPT 2:

MANUSCRIPT 3:

MANUSCRIPT 3, APPENDIX:

List of Figures

INTRODUCTION OF THE THESIS:

MANUSCRIPT *3:*

MANUSCRIPT *4:*

Acknowledgments

I wish to begin to thank my first supervisor, Prof. Dr. Sarah Teige-Mocigemba. Thank you, Sarah, for all your support throughout this dissertation process. I am grateful for what I was able to learn from your knowledge and under your supervision. Thank you for being the person you are, who cares for her employees even beyond the working hours, for the supportive surprise measures in difficult phases (e.g., anti-stress sweets or comforting flowers on my desk). Thank you for making this dissertation possible!

I further wish to thank Prof. Dr. Vincent Yzerbyt. I am very glad that we met at the ESCON 2019 in Bordeaux and for the possibility of having a research stay at your lab in Louvain-la-Neuve. During my lab visit and beyond, it was a pleasure learning from your knowledge. Thank you for your motivating and constructive feedbacks which have been encouraging for me to learn more.

Thank you, Prof. Dr. Rainer Schwarting and Prof. Dr. Christine Pané-Farré for accepting being part of the defence committee.

I wish to thank the RTG 2271 “Breaking Expectations” funded by the DFG for their generous financial support, and more importantly, the RTG 2271 network for the fruitful exchanges. I further wish to thank Anna Thorwart for her inspiring feedback on a poster presentation as part of the second RTG retreat on an earlier stage of this work. In addition, I wish to thank Ulrich Wagner for his time and the discussions we had. Thanks to all people who provided me with constructive and motivating feedback after conference talks, lab presentations as well as during scholarly exchanges.

I further wish to thank those supervisors who motivated me during my bachelors and masters studies and raised my interest for psychological science, among others, Jenny Voth, Timur Sevincer, Gabriele Oettingen, Jan Wacker, Guillermo Recio, Stefanie Brassen, Henning Gibbons and André Beauducel. In addition, I wish to thank those teachers back at school who have been motivating and not impairing one’s accomplishments, thank you Frau Beckmann, Frau Jess, Frau Ghomi, Frau Kaut, Frau Tafla, Herr Krumwiede, Herr Neitzel and Frau Schlubach-Rüping.

Thanks to all my colleagues and friends who supported me during the last years. Among others, thank you, Iris and Chris for making my start in Marburg easier. Thank you, Iris, for all the motivating voice messages specially during the last months. Thank you, Sofia, for being the companion you have been, it was more than fun visiting the balancing

courses with you. Thank you, Kiara, for being the colleague you have been and for supporting me during the final line. Thank you *Arbeitseinheit Psychologische Diagnostik* for the exchanges, lab meetings, lunch breaks, fun exam observations, and the active get togethers. Thank you, Adrian and Carmen and all members of our joint diagnostics and social psychology colloquium for the fruitful exchanges. Thank you Philipp Süssenbach for the insightful discussions that we had. Thank you, Sascha, David and Gunnar for supporting me. Thank you, Daniel and Flo for the interesting discussions that we had and the support I received.

Thank you, Sylvie Graf and Michal Bilewicz for the very interesting intergroup relations workshop at the EASP Summer School, thanks for the thrilling discussions and your motivating feedback! Thank you, Rosandra and Chiara for being such good companies at the summer school!

Thank you, Louvain Social Psychology Lab for being such a welcoming and inspiring lab. Thank you, Julie and Mathias for making my staying in Louvain-la-Neuve pleasant and for the scholarly exchange. Thank you, Emma, Marine, Julien, Matthias, Tina, Pascaline, and Antoine for being the welcoming colleagues that you have been.

I wish to further thank all the student assistants who supported my work, especially Sophie Gosch, Lea Teutenberg, Helena Huber, Marion Ihls, Hannah Mackenrodt, Lara Neudert, Lilly Müller, Bettina Salger and Ronja Haake. A big thanks targets all my participants who participated in our experiments. Thanks to Violetta Biermann, for being such a great intern and later colleague.

I would like to further thank my mentors who supported me patiently during this phase, thank you for your friendly ear.

Thank you, Erika, for the calming art you taught me, which has been a very nice complement to my PhD life.

I wish to thank my friends Sibel, Miriam, Martina, Zeynep K., Elen, Jan, Zeynep C-K., Nour, Ana, Kim, Rukiye, Sina, Nona and Pauline. Spending time with you, the phone calls and (voice) messages were heart-warming!

Thank you for being the supportive friend that you are Neda, whom I can rely on. I very much appreciate having you and I am glad that we met in Marburg! Thank you, Nivo, for being such a kind ally, I am very happy that we met in Marburg!

My special thanks targets Lilly, without your support this would not have been possible neither. Thank you for all the phone calls, long and comforting voice messages, surprise packages and meetings every now and then when we are both in the same city, which so much helped me throughout different phases. Thanks for the friendship I so much value. Thanks for being there for me when I needed you. Thank you as well for proofreading this dissertation.

Thank you, Mrs. Petersitzke! Thank you for this precious connection during the past years and for your clear view on things. Thank you for proofreading my previous works and now this dissertation, which you really wanted to see. I appreciate having you in my life.

The following thanks are expressed in my mother tongue and targets my family:

تشکر ویژه از پدرم، مادرم و برادرم به خاطر محبت و حمایت بی پایان شان. شما همیشه دانش من را در اولویت قرار دادید و بسیار حمایت کننده بودید. از همه حمایت ها و محبت های شما سپاسگزارم. مرسی مامان که اینقدر صبورانه به من گوش میدهی و ممنون بابت همه عشقی که به من میبخشی.

مایلم همچنین از پدربزرگ و مادربزرگم تشکر کنم. من خیلی خوش شانس هستم که شما را به عنوان پدربزرگ و مادربزرگم دارم. ممنون مامان آفاق عزیزم که از بچگی داستانی برام تعریف می کردی، که روزی دکتر خواهم شد. ممنونم از بابا ابی عزیزم که همیشه به من روحیه خوبی می دادید. ممنون بی بی جان عزیزم بابت این همه محبتی که با حضورت به من میدهی.

ممنون از شما دایی جواد، دایی مهرداد و عمو مجتبی که بیشتر برایم دوستان صمیمی بودید و مرا در مراحل مختلف زندگی حمایت کردید.

Finally, I wish to thank you, Jakob, for being the dearest friend and companion I could wish for. Thank you for believing in me, for motivating me and for all your support for which I am deeply thankful. Thank you for bearing the stressful final line with me and for making it easier.

For my parents, grandparents including Mrs. Petersitzke, and for Jakob.

Summary of the present work

Crimes perpetrated by migrants and asylum seekers in different European cities sparked debates about the integration of citizens and refugees from predominantly Muslim countries. Media analyses demonstrate that perpetrators' religion or cultural background are often connected with the deviant acts. However, people punish a deviant ingroup member more harshly compared to an outgroup counterpart for maintaining the positivity of the ingroup, thus expressing the so-called black sheep effect.

Considering both the literature on the black sheep effect and the stereotypes towards Muslims as stigmatized outgroup, we tested several characteristics affecting the evaluation of ingroup deviance relying on the coping with ingroup deviance model and beyond. More importantly, we provide empirical tests and results beyond the coping with deviance model, while taking into account victim's ethnicity as well as group characteristics, and shed light on differential patterns on not only individual deviance level but on cultural level.

We provided participants with alleged newspaper articles and asked them to evaluate perpetrators, victims and their cultures depending on the specific Experiment. Across eight experiments (three pre-registered experiments) with a total of N = 4642 participants (analyses sample), we operationalized designs which were complementing each other while examining the robustness of the observed patterns.

The first empirical contribution of the present dissertation examined whether German participants rely on (non)stereotypic information categories as interesting information sources to know further about. In line with the biased media representation of foreign perpetrators, participants indicated higher interest towards stereotypic information categories (e.g., religious affiliation, ethnic background) in face of an outgroup than an ingroup (German) perpetrator.

As part of the second contribution, we examined the impact of guilt certainty, crime type, and infrahumanization on perpetrator and victim blaming. We observed an interesting shift of blame: the victim was judged more harshly when the perpetrator stemmed from the outgroup. We partly observed the black sheep effect which was independent of guilt certainty. Perpetrators of sexual violence received harsher judgments than perpetrators of property crime. We further expected increased perceptions of humanness (less infrahumanization of the outgroup) coming along with equal judgments of in- and outgroup perpetrators or even outgroup discrimination. This prediction was not confirmed, however, we observed valence differences which were not predicted based on the infrahumanization theory. We further replicated the pattern of manuscript one: participants indicated higher

interest towards stereotypic information categories in face of an outgroup than ingroup perpetrator.

The third contribution of the present dissertation examined more in depth the shift of blame from the outgroup perpetrator to the ingroup perpetrator and ingroup victim in the context of sexual violence. We further used the dimensions of the stereotype content model for describing the perpetrator further beyond manipulating his ethnicity. We observed the expected black sheep effect. In addition, we observed that warm and competent perpetrators were exonerated compared to their cold and incompetent counterparts. Again, participants judged the victim more harshly when the perpetrator stemmed from the outgroup. Manipulating the victim's ethnicity indicated the same pattern: ingroup victim blaming when the perpetrator stemmed from the outgroup. Further, the ingroup victim was judged more harshly compared to the outgroup victim. More importantly, besides the judgments of individual level, here we examined attributions of blame towards the culture of the perpetrators and victims. Participants perceived the outgroup culture as more responsible for the deviant act than the ingroup culture.

As part of the fourth contribution of the present dissertation, we examined the protection of the ingroup on individual level (black sheep effect) and on cultural terms (exoneration of the ingroup culture). In addition, we tested whether these would be affected by the mere presence of the outgroup (priming), the intergroup context and the degree of the ingroup's entitativity (high vs. low). We observed robust effects on culture blaming: the ingroup culture was treated more leniently than the outgroup culture. In two out of three experiments, we observed the black sheep effect on individual level which was specifically prevalent when the intergroup context was salient. Priming and ingroup entitativity did not affect the judgments.

The present dissertation hints to a shift of blame on individual level from the outgroup perpetrator to the ingroup victim and ingroup perpetrator. However, while one outgroup individual is not judged more harshly the outgroup culture is at stake. This is to the best of our knowledge the first empirical work hinting to ingroup favoritism on both individual and cultural level which may translate to the derogation of the outgroup as a whole. In sum, we recommend differentiating on attributional levels (individual and culture) as the discrepancies observed (individual: ingroup perpetrator > outgroup perpetrator, cultural: outgroup > ingroup) have remained hidden in case of examining the judgments only on individual level. Future work may benefit from investigating further the loss of individuality of outgroup deviant members who represent a homogenous culture.

German summary of the present work: Zusammenfassung der vorliegenden Arbeit

Verbrechen, die von Migranten und Asylsuchenden in verschiedenen europäischen Städten begangen wurden, lösten Debatten über die Integration von Bürgern und Geflüchteten, überwiegend aus muslimischen Ländern, aus. Medienanalysen zeigen, dass die religiöse oder kulturelle Herkunft der Täter häufig mit den devianten Taten in Verbindung gebracht wird. Allerdings beurteilen Menschen deviante Eigengruppenmitglieder extremer als vergleichbare Fremdgruppenmitglieder, um so die Positivität der Eigengruppe aufrechtzuerhalten. Dies wird als der sogenannte *black sheep effect* bezeichnet.

Angesichts der *black sheep effect* Literatur sowie der Stereotype gegenüber Muslimen als stigmatisierter Fremdgruppe haben wir zahlreiche Charakteristika, die die Beurteilung von devianten Eigengruppenmitgliedern beeinflussen, manipuliert. Hierbei basiert die vorliegende Arbeit auf dem Modell der *coping with ingroup deviance*. Wir legen empirische Überprüfungen vor, die über das genannte Modell hinausgehen, indem wir die Ethnie der Betroffenen sowie Gruppen-Charakteristika berücksichtigen. Darüber hinaus beleuchten wir unterschiedliche Muster, nicht nur auf der individuellen, sondern auch auf der kulturellen Beurteilungsebene.

Wir haben den Versuchspersonen vermeintliche Zeitungsartikel präsentiert und sie gebeten, Täter, Betroffene und ihre Kulturen (je nach Experiment) zu bewerten. In acht Experimenten (drei prä-registrierte Experimente) mit insgesamt $N = 4642$ Versuchspersonen, die in die Analyse eingingen, haben wir einander ergänzende Designs operationalisiert, die das Testen der Robustheit der bereits beobachteten Muster ermöglicht haben.

Der erste empirische Beitrag der vorliegenden Dissertation untersuchte, ob sich deutsche Versuchspersonen auf (nicht-)stereotype Informationskategorien als interessante Informationsquellen verlassen, wenn sie weitere Informationen über einen Täter erhalten möchten. Im Einklang mit der medialen Darstellung ausländischer Täter gaben die Versuchspersonen bei einem Fremdgruppentäter ein größeres Interesse an stereotypen Informationskategorien (z. B. Religionszugehörigkeit, ethnische Herkunft) an als bei einem Eigengruppentäter (Deutscher).

Im zweiten Beitrag untersuchten wir den Einfluss von Schuldwahrscheinlichkeit, Straftat und Infrahumanisierung auf Täter- und Betroffenenurteile. Wir beobachteten eine interessante Verlagerung der Schuldzuweisung: Die Betroffene wurde stärker beschuldigt, wenn der Täter von der Fremdgruppe entstammte. Teilweise beobachteten wir den *black*

sheep effect, welcher unabhängig von der Schuldwahrscheinlichkeit war. Täter sexueller Gewalt wurden stärker verurteilt als Täter, die einen Diebstahl begangen hatten. Weiterhin erwarteten wir, dass eine zunehmende Wahrnehmung der Humanisierung der Fremdgruppe (weniger Infrahumanisierung der Fremdgruppe) zu einer gleichen Beurteilung von Eigen- und Fremdgruppentäter bzw. einer stärkeren Verurteilung des Fremdgruppentäters führen sollte. Diese Vorhersage konnte nicht bestätigt werden, allerdings beobachteten wir Valenzunterschiede, die auf Basis der Infrahumanisierungstheorie nicht vorhersagbar waren. Weiterhin replizierten wir das Muster von Manuskript 1: Versuchspersonen zeigten mehr Interesse gegenüber stereotypen Informationskategorien im Falle eines Fremdgruppentäters als im Falle eines Eigengruppentäters.

Der dritte Beitrag der vorliegenden Dissertation befasste sich eingehender mit der Schuldverschiebung vom Fremdgruppentäter zum Eigengruppentäter und Eigengruppenbetroffenen im Kontext sexueller Gewalt. Wir haben die beiden Dimensionen des *Stereotype Content Model* verwendet, um den Täter über die Manipulation seiner ethnischen Zugehörigkeit hinaus zu beschreiben. Es ließ sich der *black sheep effect* finden. Darüber hinaus haben wir beobachtet, dass warm und kompetent beschriebene im Vergleich zu als kalt und inkompetent beschriebenen Tätern entlastet wurden. Auch hier verurteilten die Versuchspersonen die Betroffene stärker, wenn der Täter von der Fremdgruppe stammte. Die Manipulation der ethnischen Zugehörigkeit der Betroffenen zeigte das gleiche Muster: eine stärkere Eigengruppen-Betroffenenverurteilung, wenn der Täter von der Fremdgruppe stammte. Darüber hinaus wurde die Betroffene der Eigengruppe im Vergleich zu der Betroffenen der Fremdgruppe stärker verurteilt. Wichtiger noch, neben den Urteilen auf individueller Ebene, untersuchten wir hier Schuldzuschreibungen bezüglich der Kultur der Täter und der Betroffenen. Die Versuchspersonen empfanden die Kultur der Fremdgruppe als stärker verantwortlich für das deviante Verhalten als die Kultur der Eigengruppe.

Im Rahmen des vierten Beitrags der vorliegenden Dissertation haben wir die Sicherung der Positivität der Eigengruppe auf individueller Ebene (*black sheep effect*) und auf kultureller Ebene (Entlastung der Eigengruppenkultur) untersucht. Darüber hinaus haben wir getestet, ob durch die bloße Anwesenheit der Fremdgruppe (Priming), den Intergruppenkontext und den Grad der Eigengruppen-Entitativität (hoch vs. niedrig) die Urteile beeinflusst werden. Wir beobachteten robuste Effekte auf kultureller Verurteilungsebene: Die Eigengruppenkultur wurde nachsichtiger behandelt als die Fremdgruppenkultur. In zwei von drei Experimenten replizierten wir den *black sheep effect*

auf individueller Ebene, der besonders ausgeprägt war, wenn der Intergruppenkontext salient war. Priming und Eigengruppen-Entitativität hatten keinen Einfluss auf die Urteile.

Die vorliegende Dissertation deutet auf eine Schuldverschiebung auf individueller Ebene vom Fremdgruppen-Täter zum Eigengruppen-Betroffenen und Eigengruppen-Täter hin. Während ein deviantes Individuum der Fremdgruppe nicht stärker verurteilt wird, steht die Kultur der Fremdgruppe auf dem Spiel. Dies ist nach unserem besten Wissen die erste empirische Arbeit, die sowohl auf individueller als auch auf kultureller Ebene auf Eigengruppen-Bevorzugung hinweist, was zu einer Diskriminierung der Außengruppe als Ganzer führen könnte. Zusammenfassend empfehlen wir, Attributionsebenen (Individuum und Kultur) zu differenzieren, denn die vorliegenden Unterschiede (individuell: Eigengruppen-Täter > Fremdgruppen-Täter, kulturell: Fremdgruppe > Eigengruppe) wären bei der Untersuchung der Urteile nur auf individueller Ebene verborgen geblieben. Zukünftige Arbeiten könnten von der weiteren Untersuchung des Verlusts der Individualität von devianten Fremdgruppenmitgliedern, die eine homogene Kultur repräsentieren, profitieren.

Introduction

"Who am I? I am in the media more than Donald Trump and his Tweets, Erdogan and his democracy or Putin and his policies. (…) For many citizens of this country, I am a key issue that needs to be resolved, because I'm a bigger threat than old-age poverty, domestic and child abuse, pollution, drug addiction, climate change or the shortage of carers and teachers. I am the one who always feels responsible for other people's mistakes, people I don't even know. When something has once again happened somewhere, I am the one who is always too embarrassed to say hello to my neighbours. I am liable for every single person's mistakes and feel threatened by every media report." (Gouma, 2019, as cited in Gümüsay, 2022 p. 56-57)

Maybe the reader already guessed to whom the writer of this letter is referring. If not, the next paragraph is solving this riddle:

"Did you recognize me? I am the refugees! (…) And all refugees. I am not a doctor, not a lawyer, neither a farmer nor a journalist, not an artist, not a salesman, neither a taxi driver nor a teacher, but the refugees. Although I come from a small town in Syria and the people in Damascus were already strange to me, since I have been in Europe, I have been one of hundreds of thousands of refugees from Syria, Pakistan, Afghanistan, Iraq, Iran and Africa. Even though we speak different languages, have different religions and pasts, let alone worldviews and opinions. But who cares about such differences, in the end we are all the refugees. Because of the war I lost friends and relatives, my apartment, job, car, my past and my homeland. But a loss that I only felt later is my individuality, which I left behind on the dinghy at the borders of Europe."

This letter, addressed to one German newspaper by Syrian lawyer Vinda Gouma (Gouma, 2019), demonstrates the scope of the present dissertation, namely, the loss of individuality for outgroups who represent others and are made responsible for other deviants' acts.

As Vinda Gouma highlighted in her letter, there has been an increase in providing a picture of a homogenous threatening outgroup as part of the media (Stürmer et al., 2019; Wigger et al., 2022) who elicits different threat types such as safety threat (Landmann et al., 2019) and on some occasions is depicted in a subtle dehumanized, that is, less human way (Bleiker et al., 2013; N. Haslam, 2006; Siem et al., 2017; Wigger, 2019).

When we look at the media analyses, we observe a bias in representation. The religion of a perpetrator (i.e., Islam) marks the largest predictor of news coverage based on an analysis conducted for the United States for the years 2006 and 2015 (Kearns et al., 2019).

Although for some perpetrators, one deviant act is attributed to mental health frames (Duxbury et al., 2018; Noor et al., 2019), for others the attribution may target terroristic motives or the deviant act is associated with their alleged homogenous (outgroup) background (Kauff, 2022; Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022). Alongside these representational biases the question arises how we react to deviant behaviour when the harm is committed by an ingroup perpetrator, as the ingroup is psychologically primary (Yzerbyt et al., 2000).

In case of coping with ingroup deviance one may choose an appropriate coping mechanism (Panitz et al., 2021; Pinquart et al., 2021) for preventing an impairment of one's social image (Tajfel & Turner, 1979). This may be expressed in more extreme evaluations of negative ingroup deviants (on individual level) compared to similar outgroup counterparts, the so called black sheep effect (Abrams et al., 2000, 2002; Marques et al., 1988; Marques & Paez, 1994; Marques & Yzerbyt, 1988). On other occasions, one may not evaluate ingroup deviants more harshly and a positive bias may emerge (Otten & Gordijn, 2014) which is expressed in more lenient judgments of the ingroup deviant compared to an outgroup deviant.

Otten and Gordijn (2014) summarized three groups of moderators (elaborated further below) which impact the coping with ingroup deviance, see Figure 1, panel A. The two reactions, a positive or a negative bias, strive to preserve a positive social identity (Ellemers & Haslam, 2012; Otten & Gordijn, 2014; Tajfel & Turner, 1979). As part of the present work, we shed light on some characteristics affecting coping with deviance when the intergroup context is salient. Furthermore, we contribute to the coping with ingroup deviance model and the black sheep effect literature by examining not only judgments on individual level but also on group (here: cultural) level. To my knowledge, no work has so far investigated the protection of the ingroup on both individual and on group level, the present dissertation fills this gap. In the next section I provide an overview of the aims of each manuscript of the present dissertation.

The present dissertation and its aims

The present work investigates discrepancies in judgments for deviant behaviour (manuscript 1-2: theft and rape, manuscript 1-4: rape). The first empirical contribution of the present work investigates whether representational biases were reflected in the information search of participants. That is, we investigated whether participants differ in their interest towards (non)stereotypic information categories about an in- and outgroup

perpetrator. Second, we shed light on the impact of some novel characteristics (manuscripts 2-4) on the evaluation of perpetrators and victims in the context of deviant acts. These characteristics include among others the social group of the perpetrator[1] and victim and the description of the perpetrator as positive vs. negative based on warmth and competence (manuscript 3) which I illustrate in the following sections (see Figure 1, panel B).

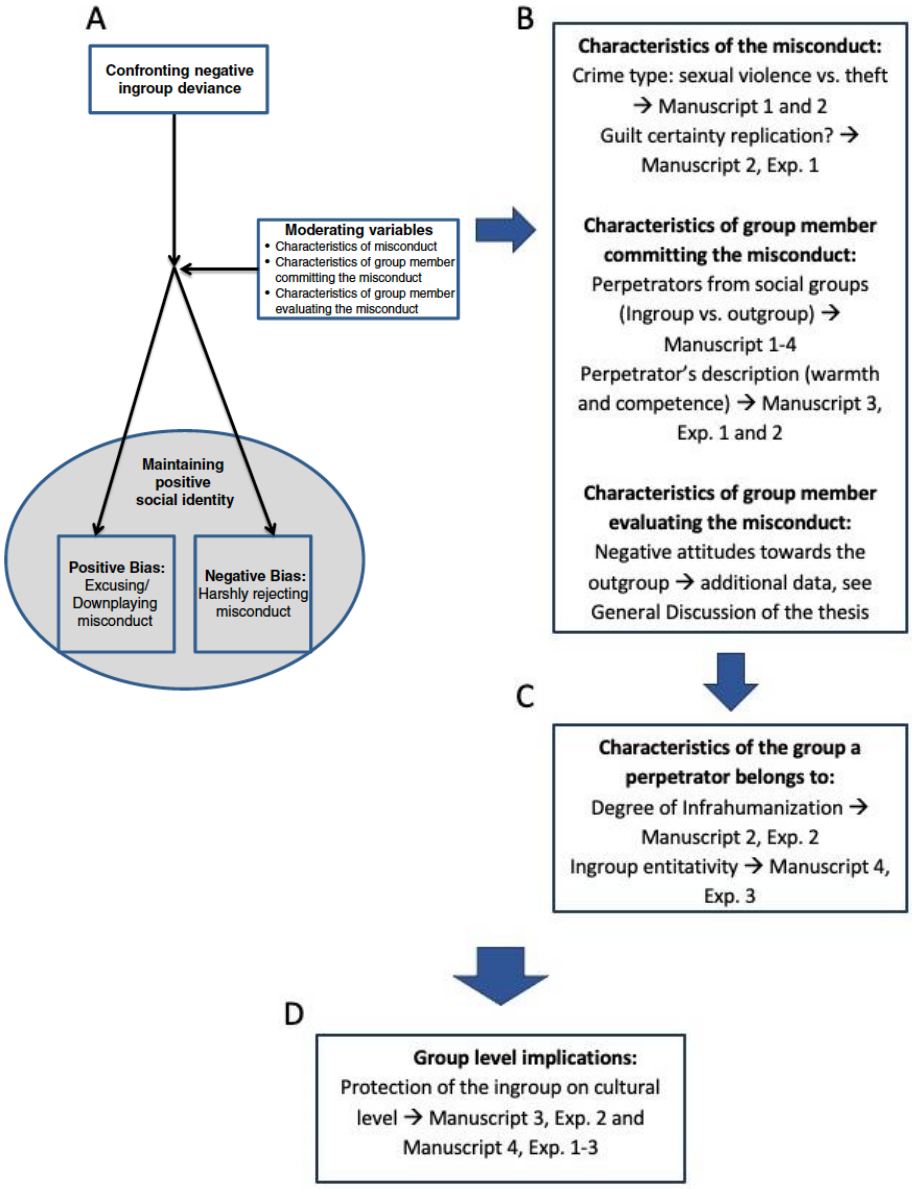
Third, and as mentioned above, the present work goes beyond the framework of the ingroup deviance model by examining characteristics of the group a member belongs to, such as degree of infrahumanization (manuscript 2, Exp. 2) and ingroup entitativity (manuscript 4, Exp. 3), see panel C. Furthermore, the present dissertation sheds light on judgments on cultural level. More precisely, we investigated how participants judged a perpetrator (and victim) on individual level and in addition in how far participants judged the responsibility of the whole culture (including norms and values) for a deviant act.[2] This group implications are complemented as part of Figure 1, panel D and examined as part of manuscript 3, Exp. 2 and manuscript 4, Exp. 1-3.

In the following sections, I will first introduce the black sheep effect and follow with the introduction of the coping with deviance model. I will discuss each of the characteristics which are summarized as part of Figure 1 before presenting the questions addressed as part of each of the manuscripts in more detail.

[1] In the following I will talk about in- and outgroup perpetrator. For the present work the ingroup includes German participants who a) had German as mother tongue and b) their parents did not have a mother tongue from a predominantly Muslim country. We did not intend to discriminate against any group, we used this exclusion criteria as the outgroup perpetrator is presented as a perpetrator from a predominantly Muslim country, Afghan or Syrian. As such we tried to collect data from an ingroup which is then confronted with an in- and / or outgroup perpetrator.

[2] It is further important to note that the present work focuses on third-party evaluations, that is the participants are not directly involved as the target of the crime, but they evaluate the harm doer (perpetrator) and the target of the crime (victim), the latter only as part of the specific works including both parties.

Figure 1: Coping with ingroup deviance model and the contribution of the present work



Notes. Panel A: Coping with ingroup deviance model (Otten & Gordijn, 2014, p. 168), panel B: overview of the characteristics manipulated as part of the present work, panel C: group level characteristics which are examined as part of the present dissertation, panel D: group level implications beyond the individual level.

The black sheep effect

We already saw as part of the introduction of this work, that one of the possible reactions towards an ingroup deviant lies in harsher judgments of anti-normative (vs. pro-normative) ingroup members which keeps the validity of ingroup norms intact (Abrams et al., 2000, 2002). This sophisticated form of ingroup favoritism (Marques et al., 1988; Marques & Paez, 1994; Marques & Yzerbyt, 1988; Yzerbyt et al., 2000) may seem contradictory as an ingroup protection strategy but it serves to preserve the positivity of the ingroup (Marques, 1990; Marques & Yzerbyt, 1988). The black sheep effect comprises the joint occurrence of ingroup bias and ingroup derogation (Marques & Yzerbyt, 1988 p. 288) as unlikeable (likeable) ingroup members are evaluated more negatively (positively) than unlikeable (likeable) outgroup members (Marques et al., 1988; Marques & Yzerbyt, 1988). For instance, the evaluation of a poor (good) speech revealed the black sheep effect in inter- and intragroup contexts (Marques & Yzerbyt, 1988). Group membership and speech performance were manipulated both between- and within-subjects as part of two experiments and the authors observed that the black sheep effect is associated with the positivity of the whole ingroup (Marques & Yzerbyt, 1988). The black sheeps are perceived as not "real members" (for a review on intergroup relations, see Yzerbyt & Demoulin, 2010). Beyond motivational explanations (i.e., social identity theory, Tajfel & Turner, 1979) there are other processes at hand for the black sheep effect to occur (Reese, 2013). Cognitive processes explaining the black sheep effect include for instance the violation of positive standards or expectancies (Biernat et al., 1999), which may be in line with a subtyping process (Carnaghi & Yzerbyt, 2007). Other discussed processes include the necessity of distancing oneself from the ingroup (Eidelman & Biernat, 2003) and different information processing strategies (Reese et al., 2013). For understanding when a positive or negative bias (the black sheep effect) in face of ingroup deviance occurs, I will depict an overview from the coping with ingroup deviance model in the next section.

Coping with ingroup deviance

Based on their empirical review, Otten and Gordijn (2014) identified three groups of moderators affecting the two reactions in face of ingroup deviance. These include the characteristics of the misconduct, the characteristics of the evaluator examining the misconduct as well as the characteristics of the group member committing the misconduct (depicted in Figure 1, panel A).

One of the characteristics of the misconduct comprises guilt ambiguity (Otten & Gordijn, 2014; van Prooijen, 2006). Feelings of anger and hostility have been observed to increase towards the ingroup deviant when the guilt of the perpetrator is certain. In case of guilt uncertainty, the so called benefit of the doubt occurred, leading to less harsher judgments of the ingroup deviant (van Prooijen, 2006). As part of manuscript 2, Exp. 1-2 we manipulated guilt certainty based on van Prooijen's (2006) work while also manipulating crime type, as another characteristic of the misconduct. We aimed to replicate the interaction effect observed by van Prooijen (2006), that is, the black sheep (dis)appearing when the ingroup perpetrator's guilt is (not) certain. The role of guilt certainty is also discussed as part of manuscript 4. Regarding crime type, there are differences in how far a perpetrator or a victim is judged more harshly (Bieneck & Krahé, 2011; Brems & Wagner, 1994; but see Reich et al., 2022). We aimed to examine the interaction between perpetrator's ethnicity and crime type as part of manuscript 2, Exp. 1.

As part of the characteristics of the perpetrator, Otten and Gordijn (2014) among others present power / leadership status. For instance, harsher reactions towards leader's anti-normative behavior was observed for severe acts whereas minor deviant acts elicited more lenient reactions, hinting to an interaction of leadership status and severity of the harm (Karelaia & Keck, 2013; Otten & Gordijn, 2014). As part of the present set of experiments, we used social groups (using names and nationalities), such as a highly stigmatized outgroup (manuscripts 1-4). We further manipulated the description of the perpetrator by using the dimensions of the stereotype content model (for a recent review, see Abele et al., 2021; Fiske, 2018; Fiske et al., 2002), see manuscript 3.

As part of the characteristics of the person evaluating the harm, social identification strength is mentioned (Otten & Gordijn, 2014). Higher ingroup identification came along with harsher evaluations of deviant acts where the hostile intent was unambiguous (unpublished manuscript by Braun et al., 2009, as cited in Otten & Gordijn, 2014). [3]

[3] Other work illustrated an inversed u-shape (quadratic) relationship between social identification and collective guilt, that is, participants with low and high levels of ingroup centrality (aspect of social identity) experienced less collective guilt than those moderate on ingroup centrality (Masson & Barth, 2019). As part of the present work, we did not find any impact of social identity on the judgments. As mentioned above there are cognitive processes which are also discussed as underlying processes of the black sheep effect.

Further, we examined characteristics which may be complemented as characteristics of a whole group. These group level characteristics include for instance infrahumanization of an outgroup (detailed further below). As part of manuscript 2, Exp. 2. we examined whether infrahumanization (the degree of humanness based on ascribed secondary emotions) affects the judgments of participants on individual level. We further looked at ingroup entitativity as another relevant group characteristic (manuscript 4, Exp. 3), as high ingroup entitativity is associated with high ingroup identification (Castano et al., 2003), we examined its impact in intergroup judgments.

In the next section, I will depict our choice of outgroup category by illustrating the way the stigmatized outgroup is depicted as part of the media (Kearns et al., 2019; Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022). Furthermore, I will conclude the next section with illustrating where the contribution of the first manuscript of this thesis lies.

The media representation of refugees, foreigners, and Muslims in the context of deviant acts

"We shouldn't forget that the people who are coming here grew up in a different religion and represent a completely different culture. Most are not Christian, but Muslim.... That is an important question, because Europe and European culture have Christian roots" – Viktor Orbán, Hungarian Prime Minister (Chadwick, 2015; Noack, 2015)

The long title of this section including three groups (e.g., refugees, foreigners, and Muslims) should not imply their relationship (not every refugee is a Muslim or vice versa). However, previous work point to a synonymous although wrong use of such constructs indicating a perception of these groups as a homogenous conglomerate (see footnote 17 in N. Kteily et al., 2015 for the association between the categories Arabs and Muslims in the United States, p. 42; Shooman, 2012). For instance, Shooman (2012) refers to the racialization of Muslims and Islam and points to a synonymous use of different group of people such as Turks, Arabs, migrants, and Muslims.

Although one out of three women worldwide have experienced physical or sexual violence caused by intimate partner violence as the most common form of violence in this area (*Intimate Partner and Sexual Violence against Women*, 2019; *RKI - GBE Berichte - Gesundheitliche Lage Der Frauen in Deutschland*, 2022; *Violence against Women*, 2021) previous media analysis, indicated a racialization of Islam as part of the attribution of sexual violence to male Muslim migrants in Germany (Stürmer et al., 2019; Wigger, 2019).

To give an example, an analysis of 163 online newspaper articles from the six most popular German online newspapers illustrated that approximately 80 percent of the examined articles referred to the suspect's ethnical background and pointed to a Muslim culture as an explanation for the sexual assaults in Cologne 2015/2016 (Stürmer et al., 2019).

When we look at attitudes towards Muslims, we observe a negativity. In most European countries (especially Eastern and central Europe), participants were more resistant towards Muslim immigration than immigration in general, with 24% of Europeans favoring a complete restriction of Muslim immigration (Gusciute et al., 2021). Women were more opposed to Muslim immigration than men which was explained by the portrayal of male Muslims as misogynistic (Gusciute et al., 2021). Several threat types (Stephan & Stephan, 1985, 2000) are associated with refugees, among them safety threat (Landmann et al., 2019). Further, previous work observed a decreased acceptance of Arab or African immigrants (Czymara & Schmidt-Catran, 2017). Also, there is a discrimination in hiring of Muslims by default (cf. Stasio et al., 2019; Unkelbach et al., 2010).

However, to my knowledge, no work has examined whether the origin of a perpetrator and his religious affiliation indeed is perceived as a valid information category. For filling this gap, we conducted a first experiment looking at possible discrepancies regarding the degree of interest German[4] participants indicate towards stereotypic information categories in face of a perpetrator from a predominantly Muslim country vs. a German perpetrator (see manuscript 1). More precisely, we examined whether participants endorse higher interest towards those information categories which are stereotypically mentioned as part of the media (e.g., religious affiliation, cultural and ethnic background) in case of a perpetrator from a predominantly Muslim country than his German counterpart.

As part of the next two paragraphs, I will provide a short overview for manuscript two by shedding light on the role of guilt certainty as well as judgments towards a victim.

Guilt certainty and the judgment of deviance

Further above, I have mentioned the pattern observed by van Prooijen (2006) which inspired manuscript 2 of the present thesis. Across four experiments van Prooijen showed the moderating effect of guilt certainty on social categorization and retributive affect. When

[4] In the present work, German participants build the ingroup. Participants with their first language as German as well as parents who were not originally from a predominantly Muslim country were included in the analyses.

guilt was certain an ingroup perpetrator was judged more harshly (participants expressed higher negative affect) but when guilt was not certain, participants expressed higher negative affect towards the outgroup perpetrator. As refugees are illustrated as part of deviant acts as part of the media, we looked at the effect of social categorization (German vs. Syrian refugee), guilt certainty and in addition crime type (rape vs. theft) as part of manuscript 2, Exp. 1. This Experiment aimed to combine the impact of factors which have not been investigated jointly so far. As part of the next paragraph, I will shortly illustrate the role of the victim as well as expected differences regarding crime type.

Coping with deviance when there is a victim

As deviant acts are sometimes targeting explicitly a victim, in the following we will briefly review literature on victim blaming. Literature on victim ethnicity exists so far for Black and White dyads (Donovan, 2007; George & Martínez, 2002; for a review on rape victim blaming see van der Bruggen & Grubb, 2014). Interracial rape victims were judged as more guilty and less credible, the act as less definitely rape (George & Martínez, 2002). In turn, the perpetrator in the interracial condition was perceived as less guilty. Relying on these results, we tested whether German victims would be judged as more guilty when harmed by an outgroup perpetrator. This would be in line with stereotypes towards male Muslims as misogynistic (Gusciute et al., 2021; Stürmer et al., 2019; Wigger, 2019). Thus, we tested whether the pattern of victim blaming would mirror those in the American context (investigated as part of manuscripts 2 and 3). We further tested whether depending on crime type the victim is judged differently. Victims of sexual assault compared to theft were judged more harshly (Bieneck & Krahé, 2011; but see Reich et al., 2022). The interacting role of crime type is also investigated as part of manuscript 2.

Turning to one of the group characteristics which we examined, namely differing ascriptions of humanness, as part of the next two paragraphs I will present an overview of the literature on differing humanness perceptions which is relevant for what we further examined as part of manuscript 2, Exp. 2.

Perceived values and the degree of humanness of outgroups

Perceiving other people as humans with comparable human essences, norms and values comes along with different perceptions and reactions towards them. Schwartz and Struch (1989) argued that the behavior towards outgroups is affected by the degree of humanity attributed to them based on the (dis-)similarity of values (hierarchy of values

between in- and outgroup). For instance, perceiving opposing values (e.g., the outgroup endorsing those values the ingroup prioritized among the least important) may elicit threats and the members of that outgroup appear as less human (Schwartz & Struch, 1989). Thus, the behavior towards the outgroup members (in a positive or antagonistic manner) is affected by the degree of similarity between the value hierarchies (Schwartz & Struch, 1989). For instance, Moss et al. (2019) examined a value violation framework which was associated with prejudices against Muslims. When US people believed that Muslims compared to Christians do not value gender equality, they reported increased prejudice towards this outgroup. Instead, when participants were told that Muslims valued gender equality (support women's rights), participants' prejudice and desire for social distance was reduced (Moss et al., 2019). This is in line with Schwartz and Struch's (1989) similarity of values eliciting threats and decreased humanness.

But how can humanness be conceptualized? The dehumanization and infrahumanization theory provide answers which I present in the next paragraph.

Dehumanization and infrahumanization

"*It is difficult to have compassion for those who lack identity and who are excluded from our community; their death does not move us in a personal way. Thus[,] when a group of people is defined entirely in terms of a category to which they belong, and when this category is excluded from the human family, then the moral restraints against killing them are more readily overcome*." (Kelman, 1973 p. 49)

The dual model of dehumanization encompasses the denial of humanness to outgroups, which comes along with two senses, one targeting human uniqueness, distinguishing humans from animals, and the other human nature, including attributes which are typically human (N. Haslam, 2006; N. Haslam & Loughnan, 2014; N. Haslam & Stratemeyer, 2016; for a review on different conceptualizations, assessments and implications of dehumanization, see N. S. Kteily & Landry, 2022). In the former case the denial is associated with the representation of the outgroup as animal-like and in the latter case as objects (N. Haslam, 2006). Blatant dehumanization has been associated with anti-refugee attitudes and behavior during the refugee crisis 2015 in European countries (Bruneau et al., 2018). Perceiving refugees as a source of threat (e.g., terrorists are trying to enter Western countries) was associated with dehumanizing immigrants and refugees (Esses et al., 2013). Looking at the literature, we perceive a dehumanization of Muslims

and Arabs by Americans as the least evolved human group followed by other outgroups (e.g., Mexican immigrants, South Koreans and Chinese people) (N. Kteily et al., 2015).

A subtle form of dehumanization, the infrahumanization theory, distinguishes between the emotions which are preserved for the ingroup (Leyens et al., 2000, 2001). People tend to reserve the so-called secondary emotions (which are unique to humans in contrast to the primary emotions which are also experienced by animals) for their ingroup, thereby infrahumanizing outgroups. Regardless of their valence, positive secondary emotions (such as hope) as well as negative ones (such as melancholy) are more likely ascribed to one's ingroup compared to an outgroup (Haslam, 2006; Leyens et al., 2000, 2001). Both blatant dehumanization and infrahumanization predict support for reducing Arab immigration (N. Kteily et al., 2015; N. S. Kteily & Landry, 2022).

In the context of violence, infrahumanization has a justification function of the ingroup's harm perpetrated against an outgroup victim (Castano & Giner-Sorolla, 2006). When the ingroup was at risk of being collectively responsible for an outgroup mass killing, participants infrahumanized the outgroup victims (Castano & Giner-Sorolla, 2006). As such, principles of morality diminish to the degree of dehumanization of the victims (Kelman, 1973).

But when we go beyond the implications of infrahuman victims, the question arises how perpetrators are perceived in the context of dehumanization and infrahumanization framework (Bastian et al., 2013; Viki et al., 2013). In the context of sexual violence, Viki et al. (2012) investigated the role of dehumanization on the rehabilitation of sex offenders. Dehumanizing sexual offenders predicted reduced support for their rehabilitation, higher sentence lengths and the social exclusion of the offender (Viki et al., 2012). However, so far, no work examined the impact of decreased infrahumanization of a stigmatized outgroup on judgments of a perpetrator. As part of manuscript 2, Exp. 2, we examined whether perceiving the outgroup as more human (we aimed to reduce infrahumanization with an experimental manipulation) would come along with less discrepancy or in other words equal judgement of the in- and outgroup perpetrators.

Another contribution to the humanness concept (N. Haslam & Loughnan, 2014), comes from one of the influential models describing the way people perceive other groups of people is the Stereotype Content Model, which I will shortly introduce in the following as it builds the framework of manuscript 3 of this thesis.

Stereotype Content Model (SCM)

The SCM postulates that social evaluations rely on two dimensions: warmth and competence (for a recent review, see Abele et al., 2021; Fiske, 2018; Fiske et al., 2002). The warmth dimension (friendliness, trustworthiness) indicates intent and the competence dimension (capability, assertiveness) indicates the ability to act in line with the intent (Abele et al., 2021; Fiske et al., 2002). The evaluations of groups as high or low on both dimensions or high in one and low in the other dimension (Yzerbyt et al., 2005) predict in turn emotional prejudices (Fiske et al., 2002). Ingroup members are perceived as warm and competent whereas Arabs or Muslims are perceived as low in both dimensions (Cuddy et al., 2009; but see Kotzur et al., 2019). It has been shown that dehumanization targets those groups falling into the low warmth and low competence dimension (Harris & Fiske, 2006). That is, for extreme outgroups (hostile and incompetent, such as homeless people or drug addicts etc.), the activation in the medial prefrontal cortex (mPFC) which is necessary for social cognition, was relatively absent whereas activation in regions associated with disgust (insula and amygdala) were observed (Harris & Fiske, 2006).

As the categorization of people does not take place on a single dimension (Cuddy et al., 2009), we used the dimensions of the SCM for further describing a perpetrator in more detail. As such, next to the impact of ethnicity, we examined the effect of his (positive, negative or no further information) description on the judgments (see manuscript 3). Previous work hint at the impact of social status of perpetrators and their evaluation. For instance, perpetrators of an armed robbery were exonerated when they belonged to the middle (vs. low) socioeconomic class (Gleason & Harris, 1976). With the manipulation of the warmth and competence of perpetrators we wanted to shed light on the question whether exonerating patterns will be observed in the context of sexual violence and whether they interact with the perpetrator's (manuscript 2, Exp. 1-2) and victim's (manuscript 2, Exp. 2) ethnicity.

As mentioned above, with the present work, we shed light on judgments beyond individual (perpetrator or victim) level. More precisely, we further examined judgments on cultural level, this brings us to the next section, where I will depict what we know from work on collective responsibility and elaborate where our contribution fits in.

From individual guilt to collective responsibility

It is observed that to the degree to which group members are perceived as interchangeable comes along with group level implications (Crawford et al., 2002; Kardos

et al., 2019; Lickel et al., 2006). This brings us to the concept of collective guilt and collective responsibility.

Group level conclusions based on individual behavior are investigated as part of the literature on collective responsibility or vicarious retribution (Lickel et al., 2003, 2006). Vicarious retribution points to aggressive reactions by non-involved ingroup members towards non-involved outgroup members who are perceived as responsible for the aggressive act of their fellow outgroup members (Lickel et al., 2006). One of the factors affecting vicarious retribution or collective blame lies in the intuitive inference that other group members are causally related to one individual's deviant act which paves the path for collective blame (Lickel et al., 2006). Interestingly, collective responsibility perceptions have been shown to guide individual behavior: Participants behave more prosocial (even on their own personal costs) when the ingroup can be perceived as collectively responsible based on their behavior (Kardos et al., 2019). That is, there is the awareness of individual behavior triggering collective responsibility and disreputation over the whole ingroup (Kardos et al., 2019).

The first investigation of intergroup attribution bias in context of international terrorism revealed that Muslim participants perceive Muslim perpetrators as black sheeps while non-Muslim participants perceived the Muslim perpetrators as "*white sheeps*" (typical for their group), leading to perceptions of higher responsibility ratings towards the Islamic group as a whole (Doosje et al., 2007). Here, attributions of responsibility were assessed with two explicit items targeting the extent participants perceived the Islamic world as responsible for terrorist attacks.

As part of manuscript 3, Exp 2 and manuscript 4, we investigated both judgments on individual and cultural level. On individual level, we aimed to test whether the black sheep effect is observed. On cultural level, we aimed to test whether the ingroup culture is exonerated compared to the outgroup culture considering what I illustrated regarding the outgroup as a homogenous threat further above. If participants would indeed judge one ingroup perpetrator more harshly than his outgroup counterpart, we would observe ingroup protection on individual level (distancing strategy from the deviant exemplar). We further expected participants perceiving the culture of the outgroup as more responsible for the deviant behavior and thus protecting the ingroup on group level.

As part of manuscript 4, we additionally examined some boundary conditions for ingroup favoritism to occur which I will depict in the following.

The role of the outgroup being (not) salient and the comparative context

Perceiving the outgroup as a threat (e.g., Landmann et al., 2019) via the media representation (Ahmed & Matthes, 2017; Hestermann, 2019; Kearns et al., 2019; Saleem et al., 2017; Soral et al., 2020; Stürmer et al., 2019; Wigger, 2019) may motivate harsher distancing strategies from these *typical* deviants. Taking into account prevalent anti-Muslim attitudes (Ogan et al., 2014; Zick, 2017; Zick et al., 2011), the assumption may arise that deviant acts such as sexual violence are mostly committed by Muslim perpetrators or refugees (Stürmer et al., 2019) but not by ingroup members. The confrontation with an ingroup deviant member may in turn violate this assumption and as such lead to a harsher black sheep effect. As such one may ask whether the mere presence of the threat eliciting outgroup category may elicit harsher judgments on ingroup individual level (i.e., stronger black sheep effect in case of the outgroup being salient than when it is not salient).

As part of manuscript 4, Exp. 1, we investigated the impact of the mere outgroup salience on the judgments. Making the outgroup salient, creates an intergroup context: participants who are primed with the outgroup category before turning to the newspaper article depicting the deviant act perpetrated by an ingroup member are in an intergroup context. The impact of the comparative context has been examined as part of previous work for instance regarding the outgroup homogeneity effect (S. A. Haslam & Oakes, 1995). The outgroup homogeneity effect was observed when only one group was judged (only the ingroup or only the outgroup), in case of both being judged the perceptions revealed no differences regarding homogeneity (S. A. Haslam et al., 1995; S. A. Haslam & Oakes, 1995). As such, as part of manuscript 4, Exp. 2-3, we more explicitly examined the impact of the comparative context for the occurrence of the black sheep effect. Using a within-subjects design, that is, participants judged both in- and outgroup perpetrators and their cultures, as such we were able to examine order effects.

I began this section with the discussion of the mere presence of a stereotyped outgroup and its possible impact on the protection of the ingroup (black sheep effect). In an intergroup context, an ingroup may be perceived as less variable (Castano & Yzerbyt, 1998; Yzerbyt et al., 2000). Perceiving the ingroup as less variable may come along with a higher threat for the ingroup when there is an ingroup deviant member. This brings us to the additional factor, which we examined as part of manuscript 4, Exp. 3, which I will introduce in the next paragraph.

Entitativity and guilt attributions

Entitativity describes the degree to which a group builds a real cohesive entity (Campbell, 1958). It goes beyond mere homogeneity as the latter is only one aspect of entitativity and entitativity further includes common goals, degree of interactions and connections between group members (Agadullina & Lovakov, 2017; Campbell, 1958; Yzerbyt et al., 2000). Entitativity comes along with strengths for the ingroup, such as a privileged social status of the ingroup as well as enhanced safety and agency perceptions (Castano et al., 2003; Sacchi et al., 2009). However, an entitative outgroup is perceived as threatening, and here entitativity is associated with prejudice (Agadullina & Lovakov, 2017) and predicts group-based aggression (Vasquez et al., 2015).

It has been shown that traits from one group member can be generalized to other group members when the group's entitativity was high, thus when members were perceived as interchangeable (Crawford et al., 2002). However, as Crawford et al. (2002) discuss, people do not endorse stereotypes against the groups Crawford et al. (2002) investigated (e.g., friends). Thus, the results cannot be directly translated to social categories such as different nationalities (Crawford et al., 2002). As they conclude in their general discussion it remains an open question in how far social categories (e.g., nationalities) would be perceived as interchangeable. The third experiment of our fourth manuscript fills this gap by manipulating entitativity for ingroup (i.e., Germans). This is the first work examining the effect of ingroup entitativity (high vs. low) for a social group in the context of deviant behavior (cf. Castano et al., 2003 on social identification and entitativity for the EU context). We examined whether reading about one's social group (Germans) being high (vs. low) in entitativity preceding the confrontation with the deviant act may lead to harsher black sheep effects because of the interchangeability of the ingroup members. In Manuscript 4, Exp. 3 we examined the impact of ingroup entitativity on the judgments.

Questions addressed as part of the present work

As illustrated as part of the individual paragraphs of this theoretical introduction, the following dissertation provides empirical answers to questions which are summarized in more detail as part of Table 1. Concrete hypotheses are depicted as part of each manuscript and the specific experiments.

Table 1. Overview of the addressed questions in each manuscript.

Manuscript 1:	• When confronted with perpetrators, which information categories do participants wish to know about? • Are there differences in the degree of interest towards different (non-)stereotypical information categories (e.g., religious affiliation) in face of a perpetrator from a predominantly Muslim country vs. a German perpetrator? • Are participants aware of the stereotypicality of specific information categories?
Manuscript 2:	• Does the social categorization of a perpetrator matter in face of the crime type? Do participants endorse a black sheep effect? • Does the benefit of the doubt replicate in the context of sexual aggression with social groups (interaction between guilt certainty and perpetrator ethnicity)? • Does less infrahumanization of the outgroup predict less discrepancy between the judgments of the in - and outgroup perpetrator?
Manuscript 3:	• Is there an exonerating effect of positively described perpetrators? • Do participants endorse a black sheep effect on individual perpetrator level? • Do participants perceive the outgroup culture as more responsible for the deviant act than the ingroup culture?
Manuscript 4:	• Do participants protect the ingroup both on individual and cultural level? • Does outgroup saliency before rating the perpetrator lead to harsher judgments of the ingroup perpetrator? • Does the degree of ingroup entitativity affect the judgments?

In the following each of the manuscripts are depicted. I further provide the reader with a summary of each manuscript before presenting them in length.

Empirical work

Summary of Manuscript 1

Khosrowtaj, Z., Biermann, V., & Teige-Mocigemba, S. (2023). "He must be a foreigner" On biased information search towards in- and outgroup wrongdoers. *Manuscript Submitted*.

Implementing a 2 (perpetrator: ingroup vs. outgroup) × 2 (crime type: sexual offence vs. theft) experimental design with the first factor varying within participants and the latter between them, we examined whether German participants' degree of interest towards (non-)stereotypical information categories about a perpetrator would vary depending on his origin and the crime done. We relied this Experiment on work illustrating a representational bias of perpetrators with a foreign background (Hestermann, 2018; Stürmer et al., 2019; Wigger, 2019) as well as discrepancies in framing used as explanation for deviant acts of in- and outgroup members such as, for instance, mental illness frame (e.g., Duxbury et al., 2018; Fabregat et al., 2019). We aimed at testing the opposite perspective, namely whether participants would indeed wish to know more about stereotypical information categories (such as ethnic background and religious affiliation) which are highlighted as part of the media, in face of a perpetrator from a predominantly Muslim country than his German counterpart. For this reason, we asked participants (N = 311, n = 213 female) to imagine reading a newspaper article depicting a deviant act (first experimental manipulation between subjects: sexual offence or theft). We did not confront participants with a detailed alleged newspaper article as we wanted to make sure not priming participants and leaving room for participants' imagination and ruling out confounding factors. We started the Experiment in an abstract and open manner. We informed participants that the supposedly imagined newspaper article revealed only the age and name of the perpetrator. We did not provide the participants with any age nor name as the main aim was the investigation of the further information categories participants were interested in.

Following this, we asked participants to spontaneously name information categories they would be interested in knowing further about. This open question aimed to test whether stereotypical information categories would be named at all. After this open question, the second experimental manipulation followed: we told participants to indicate their interest towards each of nine information categories (among them ethnic background, cultural background and religious affiliation) for both the ingroup and outgroup perpetrators, respectively (within-subjects manipulation, randomized order of

presentation). Furthermore, we assessed the degree of stereotypicality associated with each of the presented information categories for both perpetrators, respectively.

Results revealed that even before presenting our set of nine information categories, a substantive number of our sample mentioned among other information categories, the origin (ethnicity) of the perpetrator. Regarding the interest towards the presented information categories, we indeed observed a biased interest towards stereotypical information categories (ethnic background and religious affiliation) in case of the perpetrator stemming from a predominantly Muslim country than a perpetrator with German origin independent of the crime committed. Interestingly, participants were aware of the stereotypicality of stereotypic information categories such as religious affiliation, ethnic background and cultural background for the outgroup perpetrator as compared to the ingroup perpetrator. The information categories health status, occupation as well as childhood and youth were perceived as more stereotypical in case of an ingroup than an outgroup perpetrator, while participants also indicated higher interest towards the childhood and youth as well as civil status in case of the ingroup perpetrator. This latter finding is in line with ultimate attribution error (Pettigrew, 1979). Information sources which resemble external explanations such as health issues are perceived as more stereotypical in case of the ingroup perpetrator. Regarding the crime type, among others, we observed participants indicating higher interest towards the religious affiliation of a perpetrator (independent of his origin, no interaction) in case of sexual offence than theft. This might be partly in line with the racialization of Islam in the context of sexual violence (Wigger, 2019) whereas it is in line with recent findings indicating that the stereotype of a non-German perpetrator include both property and violence crimes (Bolesta et al., 2022).

This first Experiment revealed that the biased media representation is not only a phenomenon limited to the depiction of foreigners as part of the media. It also reached the information search of participants, and this stereotypical information search is even expressed by a liberal German sample. The higher interest towards stereotypic information categories is in line with previous work showing a bias in asking stereotyped targets questions in a way that one's hypothesis are more likely to be confirmed than disconfirmed (Sacchi et al., 2009; Trope & Thompson, 1997). This might be a way to search for information confirming ones' stereotypes without violating them (Panitz et al., 2021; Pinquart et al., 2021).

Manuscript 1

"He must be a foreigner"

On biased information search towards in- and outgroup wrongdoers

Zahra Khosrowtaj[1], Violetta Biermann[2], Sarah Teige-Mocigemba[1]

Authors' note

[1]University of Marburg, Gutenbergstr. 18, 35037 Marburg, Germany

[2]Osnabrück University, Seminarstr. 20, 49074 Osnabrück

This research was supported by a grant from the German Research Foundation (Research Training Group 2271: "Breaking Expectations", project number: 290878970, project 11).

We wish to thank Marion Ihls, Katharina Hildebrandt, Bettina Salger, Lilly Müller, and Lara Neudert for their assistance in study preparation, data collection and data preparation. In addition, we express our gratefulness to Carmen Lienen for her recommendation to use the software Antconc and, finally, Lawrence Anthony for his fast support regarding Antconc.

Correspondence concerning this article should be addressed to Zahra Khosrowtaj. E-mail: zahra.khosrowtaj@uni-marburg.de.

Disclosure statement No potential conflict of interest was reported by the authors.

Abstract

Peaking after the sexual assaults of New Year's Eve in Cologne 2015/2016, sexual violence is often directly attributed to the perpetrator's ethnicity, culture, and religion. We investiagted whether consumers' interest in further information is influenced by the perpetrator's group membership. In a 2 (perpetrator: ingroup vs. outgroup, within-subject factor) × 2 (crime type: sexual offence vs. theft) experimental design, we examined the degree of participants' interest in and their perceptions of stereotypicality of specific information categories. In line with our hypothesis, participants were more interested in ethnic background and religious affiliation when the perpetrator stemmed from the outgroup rather than the ingroup. Practical implications of this stereotypic information seeking are discussed considering the consequences for the outgroup as a whole.

Keywords: information search, information category, Muslim representation, media bias, stereotypic information categories

"The best way to change someone's heart is to change someone's mind" - Emile Bruneau (Resnick, 2017)

Mid May 2022 - a pupil from Essen (Germany) was arrested and taken into custody due to his plans of attacking two schools in Essen. The police found materials for constructing bombs, self-made firearms, a crossbow as well as xenophobic and right-wing extremist materials such as extreme right-wing, anti-Semitic, anti-Muslim documents in his room (Upel, 2022). The interior minister of North Rhine-Westphalia (state in western Germany) declared the records of the 16-year old pupil as his "urgent call for help" and there was an attribution to his psychological problems and suicidal thought (Deutsche Presse Agentur, 2022a, 2022b; Goldmann, 2022). Now imagine that the same terror act had been planned by a young pupil stemming from a predominantly Muslim country living in Germany or a BIPoC living in the United States. Would this change the causal attribution pattern? The present work deals with this question.

In what follows, we will first review previous work highlighting the existence of double standards in the way deviant members of different groups are portrayed in newspaper reports which in turn affects our negative attitudes and expectations and fosters prejudices (Armstrong & Neuendorf, 1992; Dixon, 2006; Jacobs et al., 2019; Ogan et al., 2014; Shaver et al., 2017; Tukachinsky et al., 2015). This review illustrates a) the different framings used for representing outgroup and ingroup perpetrators in the media, b) the biased judgement and differential information gathering when confronted with stereotype (in-)congruent crimes and c) its relation to stereotype maintenance.

News coverage and attribution bias

Several studies have found systematic differences in the descriptions of deviant acts committed by ingroup versus outgroup perpetrators. An analysis of 433 news documents covering 219 events between 2013 and 2015 illustrated that shootings by White and Latino compared to Black perpetrators are more likely attributed to mental illness. White (Latino) shooters had 19 (12) times higher odds on receiving the mental illness frame than Black shooters. Further qualitative analyses for the mental illness frame highlighted the hypocrisy at a semantic explanation level for the different racial groups: White shooters were more often than Latinos and Blacks perceived as victims of society but as the only group gained from being portrayed as stemming from a good environment. In contrast to both Whites and Latinos who also gained the good character frame and were framed as having acted out of character / unexpectedly, Black shooters never gained from these frames (Duxbury et al., 2018).

Another analysis of 199 TV reports broadcasted between January and April 2019 from the most viewed television channels in Germany and 256 newspaper articles published in 2019 indicated that 34.7% of those that mention immigrants and refugees portrayed suspects of violent crimes (crimes against life, sexual self-determination and personal freedom as well as acts of brutality) (Hestermann, 2019a). Furthermore, Hesterman (2019) observed a discrepancy in the reporting of actual violent crimes: Of the 31.4% of TV reports that explicitly mentioned the suspect's origin, 28% stated a suspect's non-German background (i.e., foreigner), while only 3.4% explicitly mentioned a suspect with German background. In stark contrast, according to the police crime statistics of 2018, 69.4% of the registered perpetrators of violent crimes were German and 30.6% non-German. In newspaper articles, this discrepancy was even more evident with 41.2% articles referring to a non-German background while only 2.9% explicitly stated a suspects' German origin (Hestermann, 2019b). Such representational bias is also observed as part of an analysis for the United states where the religion of the perpetrator was the largest predictor of news coverage after controlling for target type, fatalities, and being arrested (Kearns et al., 2018).

The content analysis of 163 articles concerning the Cologne New Year's Eve incidents (2015/2016) from the six most popular German online newspaper or magazine websites shed light on the direct attribution of the sexual assaults to the suspects' Muslim culture (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022). An increase in articles problematizing male migrants and their sexuality after Cologne 2015/2016 New Year's Eve was observed while the articles highlighted male migrants as a sexual threat based on stereotypes (Wigger et al., 2022). These media analyses in different countries with differing target groups (e.g., refugees, Muslims[5], and Blacks) depict the saliency of perpetrator's origin as well as an attribution bias based on their group membership.

Outgroup stereotypes, biased information search, and social judgments

The previous section provided evidence for an existing bias in media coverage. Such biases are especially problematic since they reflect existing stereotypes, can contribute to their maintenance, and in turn affect our social perception, information processing, and decision making (Bodenhausen & Wyer, 1985; Duncan, 1976; Panitz et al., 2021; Unkelbach et al., 2008) as is reviewed in the following section.

[5] There is a racialization of Islam (Kteily et al., 2015; Shooman, 2012) even though the ethnicity should not be equalled with one's religious affiliation. Shooman (2012) hints to the synonymous use of categories such as Muslim, Turk, Arab, and Migrant.

Social judgments are based on stereotypes when people are convinced that they possess information which enables them to judge a target (Yzerbyt et al., 1994). Participants who believed that they possessed relevant information (illusion of information) about an individual target (introverted or extraverted target), were less likely using a "don't know" response for categorizing the target and expressed stereotypical answers. The sense that information has been provided activated social judgments which were based on category stereotypes even though as part of the debriefings participants indicated not having the impression of stereotyping the target. Participants relied their judgments on stereotypes when they were to believe that individuating information was made accessible although it was not (Yzerbyt et al., 1994).

Stereotypes also affect the way people gather information: not only the amount of questions but also the way the questions were formulated revealed a bias for stereotyped targets (Trope & Thompson, 1997). Participants had to distribute questions to two sets of (non)stereotyped targets. The targets were varied within participants (i.e., vegetarians and TV producer) and the issue between participants (either an issue associated with the vegetarian or with the TV producer: killing animals for their fur or opposing government censorship). Results revealed that participants address fewer questions to stereotyped targets than to non-stereotyped target. That is, when the judged attitude was the killing of animals for their fur, participants expressed more questions to the non-stereotyped target, that is the TV producer for this stereotyped issue and vice versa. Although, participants indicated more ease formulating questions for stereotyped targets, more questions were addressed to the nonstereotyped targets. The questions addressed to the stereotyped targets were asymmetric, that is the response could better confirm than disconfirm the expectancy-related attribute (Trope & Thompson, 1997).

Asymmetric questions serve to confirm one's expectations as for this questions the affirmation response increases the likelihood of confirming one's hypothesis whereas a negation does not decrease the likelihood of that hypothesis (Sacchi et al., 2012; Trope & Thompson, 1997). Furthermore, people tended to protect the ingroup representation and based on the valence of the stereotypic traits, the information search differed: more asymmetric-confirming questions were used for confirming positive compared to negative ingroup stereotypical traits whereas for negative compared to positive stereotypic outgroup traits more asymmetric-confirming questions were addressed (Sacchi et al., 2012). More symmetric information search was used for negative ingroup and positive outgroup traits. Thus, stereotypes serve as heuristics which impact new information search (Sacchi et al.,

2012). Even in face of having the opportunity of disconfirming one's stereotypes, people seek for information which is in line with one's pre-existing stereotypes and thus do not eliminate stereotyped judgments (Cameron & Trope, 2004).

Whereas dispositional explanations are provided for outgroup members' negative acts, ingroup members' negative acts are attributed to external factors such as social rules or stating it as a misunderstanding: Positive desirable acts of ingroup members, on the other hand, are often explained by internal causes (Pettigrew, 1979; Taylor & Jaggi, 1974).

The judged harshness of verdicts is influenced by the ethnicity-crime-congruency (Jones & Kaplan, 2003; Smalarz et al., 2018). Race-crime congruent cases, that is, cases where the committed crime was more stereotypically associated with a defendant from a specific race (e.g., Black defendant and auto theft vs. embezzlement) received more negative verdicts and internal attributions were made as well as less information search was conducted compared to non-congruent crimes (Jones & Kaplan, 2003). Participants were given the choice of (confirmatory and diagnostic) additional questions for an accurate verdict. These questions targeted the crime scenarios and differed based on being confirmatory (i.e., having dealership paraphernalia) or diagnostic (i.e., physical evidence such as fingerprints). White mock jurors used a diagnostic strategy (i.e., questions which allow to test for a hypothesis and alternatives) for information gathering in case of the incongruent crime of the White defendant. They used a confirmatory information-gathering strategy (i.e., hypothesis-confirming questions, a limited information search elicited by the race-crime congruency) for both Whites and Blacks with congruent crimes. For Black defendants, a confirmatory strategy was the used information gathering strategy independent of crime congruency. This is explained as part of a general guilt hypothesis where White mock jurors are more convinced by their initial assumption of the guiltiness of the Black defendants (Jones & Kaplan, 2003). If people frequently are provided with (biased) information about outgroups via news representations, they might acquire the sense of possessing sufficient information (Yzerbyt et al., 1994) for judging other outgroup members. People strive for maintenance of their generalized beliefs and an expectation update is not achieved easily (Panitz et al., 2021; Pinquart et al., 2021).

The Present research

As seen in the first section of this work, the media uses different frames for deviant acts depending on the group membership of the wrongdoer (e.g., Duxbury et al., 2018). As part of the present work, we investigated the other side of the coin: whether participants indicate higher interests towards stereotypical information categories such as ethnic

background. We further assessed the degree to which participants perceive nine information categories as stereotypic. Moreover, we examined what associations different information categories elicit depending on the defendant stemming from the in- vs. outgroup. In addition, we focused on two crime types, namely a sexual offence and theft, for differentiating possible differences as result of crime and perpetrator constellation. Previous research has, for instance, compared sexual assault and theft in other contexts (Bieneck & Krahé, 2011) as such it is not unusual comparing these two crime types. Considering the direct attribution of sexual crimes to male Muslim migrants (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022) this will be the first investigation of information search towards German and Muslim perpetrators of sexual offence compared to theft for examining possible differences.

This in-depth investigation is to the best of our knowledge the first experimental approach to examine different information categories in case of two crimes and varying perpetrator group membership. We do not provide participants with a scenario but ask participants to imagine reading a newspaper article depicting a (theft) sexual offence where the name and age of the perpetrator are mentioned. This was due to two reasons: first, this approach would allow to test what information categories participants would be interested in without having detailed information about the perpetrator. This allows to generalize the effects as participants are not restricted nor influenced by a given scenario. Second, without any further information we can rule out confounding factors priming participants or activating demand effects.

In line with work which illustrates the salience of perpetrator's origin we expected as H1) participants naming stereotypic information categories (e.g., ethnic background or nationality) as sources that they would be interested in knowing about when only the name and age of the perpetrator of a crime (sexual offence or theft) is mentioned. This first hypothesis aimed at exploratively shedding light on the most often named categories generated by participants before we presented participants with any (stereotypic) information categories.

H2-H5 relate to the list of nine information categories we presented to participants. As part of H2) we expected German participants indicating higher degrees of interest towards the categories ethnic & cultural background, and religious affiliation towards a perpetrator from a predominantly Muslim country compared to a German perpetrator.

We further expected H3) participants to perceive ethnic , cultural background and religious affiliation as more stereotypical of an outgroup perpetrator than of an ingroup perpetrator.

As part of H4) we expected religious affiliation, ethnic and cultural background being rated as more stereotypical of an outgroup sexual perpetrator than an outgroup theft perpetrator taking into account the prevalent link of sexual crimes to the ethnical background and religion of the perpetrators following the assaults in Cologne (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022).

As part of H5), we exploratively expected participants more frequently naming stereotypic associations concerning information categories targeting the outgroup than the ingroup perpetrator.

Methods

Sample size estimation

We conducted a sample size estimation using PANGEA (v0.2) web app (https://jakewestfall.shinyapps.io/pangea) for our 2 (group: German vs. perpetrator from a predominantly Muslim country) × 2 (crime: sexual offence vs. theft) design for the former factor varying within participants. We added participants as random factor, crime and group as fixed factors with two levels each. Participants were crossed with group (group factor varies within participants) and were nested in the crime factor (i.e., crime as between subjects factor). We expected a small effect size d = .30 for the interaction between perpetrator × crime. This power analysis indicated a power of .887 with a total of N = 300 participants (n = 150 per crime condition).

Participants

We collected data through several online ads and postings and offered compensation in the form of participation in a raffle for three gift cards (total value 50€), as well as compensation via course credit. The questionnaire was implemented in the web-based survey tool SoSci Survey (Leiner, 2016) and available in desktop and mobile format. In total, N = 324 participants filled out the entire questionnaire. Altogether, n = 13 had to be excluded based on our a priori defined exclusion criteria: they indicated a first language other than German (n = 8), indicated at least one parent's first language being majorly spoken in a predominantly Muslim country or one parent with an origin of a predominantly Muslim country (n = 3), did not indicate their parents' descent (n = 1), or finished the questionnaire too fast according to SoSci Survey's quality indicator for speeding (n = 1). Thus, the analysis sample included N = 311 (n = 213 female, n = 86 male, n = 8 diverse, n

= 4 did not indicate their gender) with an average age of *M* = 23 years (*SD* = 7.1 years; range: 18 – 75 years).

The majority of participants was currently enrolled at a university (*n* = 307) in a variety of different study fields (e.g., Psychology *n* = 55 including *n* = 3 economic psychology).

Overall procedure and measures

We randomly assigned participants to one of the crime conditions. First, we asked participants to imagine reading a newspaper article depicting either a sexual offence or a theft. We told participants that this imagined newspaper article included the perpetrator's name and age and asked which further information about the perpetrator they were spontaneously interested in using an open text field format. We included this first measure to assess spontaneous information categories generated by the participants before the two perpetrator groups as well as our list of information categories were introduced. More precisely, first, we told participants to imagine an article describing one of the crime types: "Please imagine that you are reading a newspaper article about a criminal offense described as a sexual offence (or theft). The article tells you the name and age of the offender who committed this crime." Next, they had to indicate spontaneously which information categories they would like to know about: "The following question relates to the article (sexual offense vs. theft) you envisioned on the previous page. What other information would you like to know about the perpetrator?".

Subsequently, participants worked through the interest items. Here we introduced the second experimental manipulation: perpetrator's origin: "The following question further targets the article (sexual offense vs. theft) you envisioned earlier. The article describes that the perpetrator comes from a predominantly Muslim country (is of German origin). In which of the following areas would you like more information about the perpetrator?"

Across participants, it was counterbalanced whether the interest items were depicted for the outgroup or the ingroup perpetrator first. We asked participants to indicate their degree of interest for each of the nine information categories (ethnic background, childhood and youth, religious affiliation, civil status, occupation, current social environment, health status, educational level and cultural background) with the following item "For which of the following categories would you like to receive more information about the perpetrator?" on a 7-point scale ranging from 1 = *not interested at all* to 7 = *absolutely interested*. Participants then worked through the same items for the perpetrator counterpart. Following this, we asked participants to indicate their associations for each of the nine categories.

Here, participants indicated what characterized the nine categories for the German perpetrator and a perpetrator from a predominantly Muslim country, respectively, again using an open text field format: “Please go through each information category and indicate what would characterize a Muslim (German) offender for each information category (e.g., what ethnic background would they have).”

Next, we assessed the degree of stereotypicality of the information categories for both perpetrators, respectively, with the following item “Are the following categories clichéd or typical of a perpetrator with German background (from a predominantly Muslim country)?” on a 7-point scale ranging from 1 = *not stereotypic at all* to 7 = *absolutely stereotypic*. After that, we asked whether there were any further information categories participants would like to receive information about, and they were given the opportunity to shortly define their entry. Finally, we collected demographic variables (e.g., gender, age, education, mother tongue, political orientation, voting behavior) and thanked participants for their participation.

Results

In the following section, we will report the results following the order of the hypotheses. We will first present, which information categories participants spontaneously generated without knowing about the perpetrator’s origin (H1), followed by the degree participants were interested in the nine information categories for the in- and outgroup perpetrator (H2). We will then present the results of the stereotypicality ratings of the nine categories (H3 - H4). Finally, we will present what frequent spontaneous associations each of the nine categories elicited for both perpetrators (H5).

Spontaneous frequently generated categories

Recall that participants were only informed about one of the two offence conditions. In a first step, we clustered the information categories that participants generated. Among these categories participants named for instance crime circumstances, sociodemographics of the perpetrator as well as ethnicity. Table 1 as part of the appendix contains the details regarding the categories which were at least mentioned by 10% of the sample in one of the two crime conditions. In line with our first explorative hypothesis, a substantive number of participants, 28.4%, named stereotypic categories related to the ethnicity of the perpetrator in the sexual offence condition as compared to 21.2% in the theft condition. This is remarkable given that we did not prime the participants with any stereotypic categories beforehand.

Degree of interest towards the information categories

Using R Studio version 1.1.453, we ran mixed model ANOVAs to test for effects of perpetrator group and crime type on interest in and stereotypicality of the information categories (H2, H4 & H5). These analyses were conducted with the ezANOVA package (v4.4.-0). We set $\alpha < .05$ for all analyses.

Recall that as part of H2 we were interested in differential patterns regarding ethnic and cultural background and religious affiliation. For this purpose, we conducted 2 × 2 ANOVAs with crime type as between subjects factor, perpetrator as within subjects factor and the interest in the information categories as dependent variable. Table 2 as part of the appendix includes all descriptive and test statistics regarding interest and stereotypicality of the information categories which are summarized in the following.

Regarding the information category ethnic background as dependent variable we observed a main effect of perpetrator, $F(1,309) = 4.44$, $p = .036$, $\eta^2_G = .001$, $\eta^2_p = .014$ such that participants were more interested in the ethnic background of the outgroup perpetrator ($M = 3.34$, $SD = 1.94$) than of the ingroup perpetrator ($M = 3.22$, $SD = 1.90$), thus supporting our second hypothesis. There was no main effect of crime type, nor an interaction effect, both p's $\geq .140$.

Regarding religious affiliation as dependent variable we again observed a significant main effect of perpetrator, $F(1,309) = 9.37$, $p = .002$, $\eta^2_G = .003$, $\eta^2_p = .029$: participants indicated higher interest in the religious affiliation of the outgroup perpetrator ($M = 3.02$, $SD = 1.93$) than of the ingroup perpetrator ($M = 2.83$, $SD = 1.77$). We further observed a significant main effect of crime type, $F(1,309) = 9.66$, $p = .002$, $\eta^2_G = .027$, $\eta^2_p = .030$ such that participants reported higher interest towards the perpetrator's religious affiliation in the sexual offence condition ($M = 3.23$, $SD = 1.87$) than in the theft condition ($M = 2.62$, $SD = 1.80$). There was no significant interaction between crime type × perpetrator, $p = .119$.

Unexpectedly, participants' degree of interest towards cultural background did not indicate any differences: There was no main or interaction effect, all p's ≥ 062.

The following explorative analysis for the remaining six information categories revealed significant patterns for interest in childhood and youth, where we observed a significant main effect of perpetrator group, $F(1,309) = 5.80$, $p = .017$, $\eta^2_G = .001$, $\eta^2_p = .018$. Participants indicated higher interest towards the childhood and youth of the ingroup

perpetrator (M = 5.19, SD = 1.84) than of the outgroup perpetrator (M = 5.07, SD = 1.82). We did not observe any other effect regarding childhood and youth, all p's ≥ .204.

Regarding the civil status we did not observe any main effect, both p's ≥ .053. Interestingly, there was a significant interaction effect between crime type × perpetrator, $F(1,309) = 4.05$, $p = .045$, $\eta^2_G = .001$, $\eta^2_p = .013$. Post-hoc pairwise comparisons revealed that participants in the sexual assault condition indicated higher interest in the civil status when the perpetrator stemmed from the ingroup (M = 4.66, SD = 1.93), $t(154) = 2.28$, p = .012, than when he stemmed from the outgroup (M = 4.46, SD = 1.99). In contrast, participants in the theft condition showed no difference regarding their interest in civil status for the ingroup and the outgroup perpetrator, $t(155) = -0.61$, $p = .54$.

Regarding the interest in occupation, current social environment, health status and educational level as dependent variables indicated no significant main or interaction effect, all p's ≥ .058[6].

Stereotypicality of the information categories

Regarding the degree of stereotypicality of the information categories (H3), we conducted another set of mixed model ANOVAs with the same predictors as before.

Participants' perception of the stereotypicality of the ethnic background indicated a main effect of perpetrator, $F(1,309) = 480.46$, $p < .001$, $\eta^2_G = .41$, $\eta^2_p = .609$, such that participants rated ethnic background as more stereotypical of the outgroup perpetrator (M = 5.78, SD = 1.47) than of the ingroup perpetrator (M = 3.10, SD = 1.70) which was in line with our prediction. We did not observe any other significant effect, both ps ≥ .54.

Regarding religious affiliation we observed a main effect of perpetrator, $F(1,309) = 392.97$, $p < .001$, $\eta^2_G = .37$, $\eta^2_p = .560$, such that participants rated religious affiliation as more stereotypic of the outgroup perpetrator (M = 5.67, SD = 1.59) than of the ingroup perpetrator (M = 3.12, SD = 1.73). No other effects were observed, both ps ≥ .089.

Regarding the stereotypicality of cultural background we observed a main effect of perpetrator, $F(1,309) = 394.06$, $p < .001$, $\eta^2_G = .34$, $\eta^2_p = .56$, such that participants rated cultural background as more stereotypic of the outgroup perpetrator (M = 5.84, SD = 1.39) than the ingroup perpetrator (M = 3.52, SD = 1.81) which again was conform with our

[6] There was a descriptive trend for higher interest towards health status for the ingroup (M = 4.93, SD = 1.84) than the outgroup perpetrator (M = 4.85, SD = 1.85), $F(1,309) = 3.62$, $p = .058$.

prediction. There was no main effect of crime type, nor an interaction effect, both *p*'s ≥ .094.

Regarding the stereotypicality of childhood and youth we observed a main effect of perpetrator, $F(1,309) = 42.88$, $p < .001$, $\eta^2_G = .043$, $\eta^2_p = .122$, such that participants viewed the childhood and youth as more stereotypic of the ingroup perpetrator ($M = 4.77$, $SD = 1.74$) than of the outgroup perpetrator ($M = 4.05$, $SD = 1.66$). There was no other significant effect, both *p*s ≥ .648.

Regarding the stereotypicality of occupation as dependent variable we observed a main effect of perpetrator, $F(1,309) = 23.03$, $p < .001$, $\eta^2_G = .02$, $\eta^2_p = .069$, such that participants rated occupation as more stereotypic of the ingroup perpetrator ($M = 4.10$, $SD = 1.76$) than of the outgroup perpetrator ($M = 3.60$, $SD = 1.66$). In addition, there was a main effect of crime type, $F(1, 309) = 14.31$, $p < .001$, $\eta^2_G = .03$, $\eta^2_p = .044$, such that the participants rated occupation as more stereotypic in the case of theft ($M = 4.15$, $SD = 1.17$) than of a sexual offence ($M = 3.55$, $SD = 1.70$). There was no interaction effect between crime type × perpetrator, $p = .202$.

Regarding the stereotypicality of health status as dependent variable we observed a main effect of perpetrator, $F(1,309) = 83.29$, $p < .001$, $\eta^2_G = .07$, $\eta^2_p = .212$, such that participants rated health status as more stereotypic of the ingroup perpetrator ($M = 4.06$, $SD = 1.96$) than of the outgroup perpetrator ($M = 3.09$, $SD = 1.61$). We did not observe any other effect, both *p*s ≥ .104.

Participants did not express any differential patterns regarding the civil status, educational level, and current social environment, all *p*s ≥ .060.

Unusually frequent associations with the information categories

Recall that participants were asked to indicate what would characterize a Muslim vs. German perpetrator for each of the nine categories after they indicated their interest towards the categories for both perpetrators and before they rated the stereotypicality of them. We used the freeware corpus and text analysis software Antconc (Anthony, 2022) for examining participants' associations for all information categories, separately for the two crime types (H5). We used the keyword list tool of Antconc that allows the detection of unusually frequent words compared to a reference corpus (*Keyword List — AntConc Manual Documentation*, n.d.). We ran the keyword analysis separately for the theft and sexual offence conditions and compared the ingroup (reference corpus) and outgroup (target corpus) conditions (18 keyword analysis for nine categories times two crime

conditions). For the keyword statistics, we used Log-Likelihood 4-term with a threshold of $p < .05$ with Bonferroni correction (Anthony, personal communication, June 01, 2022). Interestingly, for religious affiliation and cultural background, Islam / Muslim was a frequent association with a perpetrator from a predominantly Muslim country compared to the German counterpart for the theft condition. The word (type) *war* was significantly more often mentioned and associated with the outgroup compared to the ingroup perpetrator condition again for theft as crime. Using the concordance tool of Antconc software (Keyword in context), the word war is depicted in the context (phrase) indicated by participants. Some examples (English translations) were as follows: *affected by war / flight, difficult relationships, strict upbringing, shaped by war / terror, grew up in war etc., negative childhood due to war or violence*.

Discussion

With the present work, we examined whether German participants indicate higher interest towards stereotypical information categories about an outgroup perpetrator compared to his ingroup counterpart as part of two crimes. Our first explorative analysis of information categories generated by participants revealed that a substantive number of our sample in the sexual offence (theft) condition mentioned information categories concerning the ethnicity of the perpetrator as categories they were interested in receiving further information about. As such, for some participants ethnicity is perceived as a relevant information category in the context of crime related newspaper articles. It is concerning that ethnic origin was among the generated information categories following an open question, where one might expect reluctance of participants indicating stereotypical categories at all. Given possible self-presentational distortions, this number might be underestimated.

Regarding our set of information categories and in line with our prediction, we observed that participants expressed higher interest towards the ethnic background and religious affiliation of a perpetrator from a predominantly Muslim country compared to a German perpetrator, independent of the crime committed. The present work is the first empirical finding proving for higher degree of interest for ethnic background and religious

affiliation as information categories for outgroup perpetrators than for ingroup perpetrator. This was true for our even rather left politically oriented German sample[7].

In contrast to our expectation, we did not observe any difference in the degree of interest for the information category cultural background, which might be caused by different understandings of cultural background (e.g., cultural goods vs. traditional or religion-oriented culture). Even though the degree of interest towards the cultural background did not reach significance, responses on the degree of stereotypicality indicated that participants are aware of the stereotypicality of cultural background in case of an outgroup perpetrator.

Further, participants indicated higher interest towards the religious affiliation in the sexual offence condition compared to the theft condition. We did not find evidence for our expected interaction effect (perpetrator times crime); instead, participants indicated higher stereotypicality ratings for the outgroup perpetrator than for the ingroup perpetrator and, independent from the perpetrator, they were more interested in the religious affiliation in the sexual offence than theft condition. This is only partly in line with work reviewed in the introduction showing that there is a direct link between sexual crimes and the outgroup perpetrator's religion (Stürmer et al., 2019; Wigger, 2019).

This unexpectedly unobserved interaction effect between crime and perpetrator might be explained by recent results observed by Bolesta and colleagues (2022) who asked participants about their image concerning criminals. Interestingly, the stereotypes towards criminals revealed three clusters where the non-German perpetrator cluster was indicated by 50% of the sample and included both property and violence crimes as stereotyped offences (Bolesta et al., 2022). For the second biggest cluster with the German perpetrator (endorsed by 40% of the sample), no specific crime was identified. This might point to the conclusion that whereas with a German perpetrator no specific crime is associated, with the non-German perpetrator both theft and sexual offences could be associated.

Explorative analyses revealed participants to indicate more interest towards the childhood and youth of the German perpetrator than of the perpetrator from a predominantly Muslim country. This might indicate that the cause for the offence is externally attributed to the childhood of the ingroup perpetrator (cf. Pettigrew, 1979). On

[7] Most participants indicated at least rather left political orientation (85.2%) and would vote for the Green party (Alliance 90 / The Greens) if there was an election on the upcoming Sunday (42.8%).

the other hand, our corpus analysis revealed that the keyword *war* was significantly more frequently associated with the outgroup perpetrators' (vs. ingroup perpetrator's) childhood and youth in the theft condition. Although this was not the case for the perpetrator in the sexual offence condition it is striking and might point to possible associations regarding a negatively affected childhood and youth of a perpetrator from a predominantly Muslim country and this negative association might be transferred via the media representation (cf. Fabregat et al., 2019). The prevalence of the word *war* in association with the outgroup perpetrator might speak for a potential equating a refugee background for a perpetrator from predominantly Muslim country, or a homogenous perception of Muslim countries shaped by war. However, this must be interpreted with caution and other explanations might account for this finding as well.

Another explorative observation targets the information category civil status: participants indicated higher interest towards the civil status of the ingroup than the outgroup in the sexual offence condition; there was no difference between the degree of interest observed in the theft condition. One might post hoc argue that participants could be in search for external explanations for the ingroup but not the outgroup perpetrator's misconduct.

In addition, the information categories health status, occupation as well as childhood and youth were perceived as more stereotypic for the ingroup than for the outgroup perpetrator, and participants perceived occupation as more stereotypic in the theft compared to the sexual offence condition, which might explain the need to steal.

Overall, the pattern in the categories which are more interesting and stereotypic for the in- and the outgroup perpetrator are in line with the ultimate attribution error (Pettigrew, 1979). Ethnic background and religious affiliation as more internal attributes are perceived as sources which are more interesting and stereotypic when participants are confronted with an outgroup perpetrator. On the other hand, information sources such as health status, occupation, and childhood and youth, which depict external factors, are perceived as more stereotypic when participants are faced with an ingroup perpetrator. Importantly, these results were obtained although participants had the chance to indicate that they were not interested at all in particular information categories and despite the clear instructions highlighting the group membership of the perpetrators as part of a within subjects' manipulation. Furthermore, it is thought-provoking that the health status of the ingroup perpetrator is perceived as a more stereotypic information source, which we perceive as in

line with the differential medial frames used for different perpetrators (Duxbury et al., 2018; Fabregat et al., 2019).

Implications beyond one male Muslim migrant deviant

The results of this experiment hint to the biased interest towards specific information categories about perpetrators which might have severe consequences and implications beyond the individual perpetrator which we will discuss in the following. Immigrants have been found to be less visible and at the same time, more negative frames were used for representing them (Fabregat et al., 2019). An analysis of four local newspapers of Barcelona (205 articles from the years 2000 to 2012) revealed that 57% of the frames used to represent immigrants were negative and missed the chance of helping integration (Fabregat et al., 2019).

Given the racialisation and islamicisation of sexual violence (Wigger, 2019), male Muslims are more likely to be dehumanized and perceived as social threats as part of the media (Wigger et al., 2022). The more people believed the veracity of the culture explanation, the more feelings of symbolic threat (Stephan et al., 2002; Stephan & Stephan, 1996) were related to radical reactions towards refugees, that is, right-wing violence, support for armed self-defence, and support for closing Germany's borders to immigrants (Stürmer et al., 2019). When US people believed that Muslims compared to Christians do not value gender equality, they reported increased prejudice towards this outgroup (Moss et al., 2019). Instead, when Moss et al. (2019) told participants that Muslims valued gender equality, participants' prejudice and desire for social distance was reduced. Immigration supporters in Germany tended to attribute a suicide attack done by an outgroup perpetrator to mental illness rather than to terrorism as motive and this in turn predicted their being less punitive towards immigrants as a whole compared to participants opposing immigration (Noor et al., 2019).

In most European countries (especially Eastern and central Europe), participants were more resistant towards Muslim immigration than immigration in general, with 24% of Europeans favoring a complete restriction of Muslim immigration (Gusciute et al., 2021). Women were more opposed to Muslim immigration than men which was explained by the portrayal of male Muslims as misogynistic (Gusciute et al., 2021). The framing used as part of media representations thus might not only affect one single deviant but also their whole group (Doosje et al., 2007; Khosrowtaj, Teige-Mocigemba, et al., 2023; Khosrowtaj, Yzerbyt, et al., 2023; Lickel et al., 2006; Noor et al., 2019).

Using media extended (i.e., knowing about close relationships between in- and outgroup members) and parasocial (i.e., mass-mediated interpersonal interaction) contact (Fabregat et al., 2019; Schiappa et al., 2005; Wright et al., 1997) and providing unbiased knowledge about the outgroup and the religion (cf. Pettigrew & Tropp, 2008) may be promising on societal and political levels for improving the integration of immigrants and refugees as well as modifying negative attitudes towards these minorities. As religious discrimination was associated with decreased life satisfaction among Muslim American participants (Bassioni & Langrehr, 2021) it is promising to improve integration and immigrants well-being.

Conclusions

The present work is to the best of our knowledge the first experiment examining a) participants' interest towards specific information categories when confronted with an ingroup vs. outgroup perpetrator as well as b) the degree to which such information categories are perceived as stereotypical for an in- and outgroup perpetrator while considering two crimes. It is the first empirical work showing that German participants indicate higher interest towards stereotypic information categories in case of a perpetrator from a predominantly Muslim country than a German perpetrator. In contrast to other work, participants of our experiment imagined reading an article and were asked which information categories they were interested reading about and how far they rated them as stereotypic. It stands out that participants indicated higher interest towards the ethnic background and religious affiliation of the outgroup than ingroup perpetrator and judged these categories as more stereotypic of the outgroup than ingroup perpetrator. It is alarming that a perpetrator's religious affiliation seemed to be a relevant information source at all when the crime committed was sexual. This might be seen as participants' strategy seeking for information which confirmed their stereotypes, a process called assimilation (Panitz et al., 2021; Pinquart et al., 2021). Based on the results of the present work one might conclude that the biased representations depicted as part of the introduction of this work has already become consolidated in information search such that stereotypic information categories are perceived as relevant criteria to judge a perpetrator.

References

Anthony, L. (2022). *AntConc* (4.0.10) [Windows].

Armstrong, G. B., & Neuendorf, K. A. (1992). TV Entertainment, News, and Racial Perceptions of College Students. *Journal of Communication*, *42*(3), 153–176. https://doi.org/10.1111/j.1460-2466.1992.tb00804.x

Bassioni, R., & Langrehr, K. (2021). Effects of Religious Discrimination and Fear for Safety on Life Satisfaction for Muslim Americans. *Journal of Muslim Mental Health*, *15*(1), Article 1. https://doi.org/10.3998/jmmh.133

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

Bodenhausen, G. V., & Wyer, R. S. (1985). Effects of stereotypes in decision making and information-processing strategies. *Journal of Personality and Social Psychology*, *48*(2), 267–282. https://doi.org/10.1037/0022-3514.48.2.267

Bolesta, D., Oehme, & Führer, J. L. (2022). *Stereotype über Straftäter:innen* (Panel Zur Wahrnehmung von Kriminalität Und Straftäter:Innen (PaWaKS): Ergebnisse Der Ersten Erhebungswelle. Zentrum Für Kriminologische Forschung Sachsen e.V.). https://www.zkfs.de/pawaks/

Cameron, J. A., & Trope, Y. (2004). Stereotype-Biased Search and Processing of Information About Group Members. *Social Cognition*, *22*(6), 650–672. https://doi.org/10.1521/soco.22.6.650.54818

Deutsche Presse Agentur. (2022a). *Hinweise auf psychische Probleme bei Tatverdächtigem*. Süddeutsche.de. https://www.sueddeutsche.de/panorama/notfaelle-essen-hinweise-auf-psychische-probleme-bei-tatverdaechtigem-dpa.urn-newsml-dpa-com-20090101-220512-99-259764

Deutsche Presse Agentur. (2022b, May 12). Essen: Hinweise auf psychische Probleme bei Tatverdächtigem. *ZEIT ONLINE*. https://www.zeit.de/news/2022-05/12/hinweise-auf-psychische-probleme-bei-tatverdaechtigem

Dixon, T. L. (2006). Psychological Reactions to Crime News Portrayals of Black Criminals: Understanding the Moderating Roles of Prior News Viewing and Stereotype Endorsement. *Communication Monographs*, *73*(2), 162–187. https://doi.org/10.1080/03637750600690643

Doosje, B., Zebel, S., Scheermeijer, M., & Mathyi, P. (2007). Attributions of

Responsibility for Terrorist Attacks: The Role of Group Membership and Identification. *International Journal of Conflict and Violence (IJCV)*, *1*(2), Article 2. https://doi.org/10.4119/ijcv-2749

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), 590–598. https://doi.org/10.1037/0022-3514.34.4.590

Duxbury, S. W., Frizzell, L. C., & Lindsay, S. L. (2018). Mental Illness, the Media, and the Moral Politics of Mass Violence: The Role of Race in Mass Shootings Coverage. *Journal of Research in Crime and Delinquency*, *55*(6), Article 6. https://doi.org/10.1177/0022427818787225

Fabregat, E., Medina-Bravo, P., & Meyers, M. (2019). NEIGHBORHOOD PRESS AND IMMIGRATION A frame-building analysis of the depiction of immigrants. *Journalism Practice*, *13*(5), Article 5. https://doi.org/10.1080/17512786.2018.1521705

Goldmann, R. (2022, May 12). *NRW-Innenminister Reul zu Essen: Bombenanschläge verhindert*. zdf heute. https://www.zdf.de/uri/bf300004-2774-4a85-8405-36d68d05a177

Gusciute, E., Muhlau, P., & Layte, R. (2021). All Welcome Here? Attitudes towards Muslim Migrants in Europe. *International Migration*, *59*(5), 149–165. https://doi.org/10.1111/imig.12810

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

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

Jacobs, L., Boukes, M., & Vliegenthart, R. (2019). Combined Forces: Thinking and/or Feeling? How News Consumption Affects Anti-Muslim Attitudes through Perceptions and Emotions about the Economy. *Political Studies*, *67*(2), 326–347. https://doi.org/10.1177/0032321718765696

Jones, C. S., & Kaplan, M. F. (2003). The Effects of Racially Stereotypical Crimes on Juror Decision-Making and Information-Processing Strategies. *Basic and Applied Social Psychology*, *25*(1), 1–13. https://doi.org/10.1207/S15324834BASP2501_1

Kearns, E., Betus, A., & Lemieux, A. (2018). *Why Do Some Terrorist Attacks Receive More Media Attention Than Others?* (SSRN Scholarly Paper ID 2928138). Social Science Research Network. https://papers.ssrn.com/abstract=2928138

Keyword List—AntConc Manual documentation. (n.d.). Retrieved 25 May 2022, from https://antconc-manual.readthedocs.io/en/latest/keyword_list.html

Khosrowtaj, Z., Teige-Mocigemba, S., & Yzerbyt, V. (2023). "(N)One of us but all of them!" Ingroup favoritism on individual and group level in the context of deviant behavior. *Manuscript Submitted*.

Khosrowtaj, Z., Yzerbyt, V., Süssenbach, P., & Teige-Mocigemba, S. (2023). On the effect of perpetrator's description on inter- and intragroup guilt attributions. *Manuscript Submitted*.

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

Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., & Schmader, T. (2006). Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review*, *10*(4), 372–390. https://doi.org/10.1207/s15327957pspr1004_6

Moss, A. J., Blodorn, A., Van Camp, A. R., & O'Brien, L. T. (2019). Gender equality, value violations, and prejudice toward Muslims. *Group Processes & Intergroup Relations*, *22*(2), 288–301. https://doi.org/10.1177/1368430217716751

Noor, M., Kteily, N., Siem, B., & Mazziotta, A. (2019). "Terrorist" or "Mentally Ill": Motivated Biases Rooted in Partisanship Shape Attributions About Violent Actors. *Social Psychological and Personality Science*, *10*(4), 485–493. https://doi.org/10.1177/1948550618764808

Ogan, C., Willnat, L., Pennington, R., & Bashir, M. (2014). The rise of anti-Muslim prejudice: Media and Islamophobia in Europe and the United States. *International Communication Gazette*, *76*(1), 27–46. https://doi.org/10.1177/1748048513504048

Panitz, C., Endres, D., Buchholz, M., Khosrowtaj, Z., Sperl, M. F. J., Mueller, E. M., Schubö, A., Schütz, A. C., Teige-Mocigemba, S., & Pinquart, M. (2021). A Revised Framework for the Investigation of Expectation Update Versus Maintenance in the Context of Expectation Violations: The ViolEx 2.0 Model. *Frontiers in Psychology*, *12*, 5237. https://doi.org/10.3389/fpsyg.2021.726432

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, *38*(6), Article 6. https://doi.org/10.1002/ejsp.504

Pinquart, M., Rothers, A., Gollwitzer, M., Khosrowtaj, Z., Pietzsch, M., & Panitz, C. (2021). Predictors of Coping With Expectation Violation: An Integrative Review. *Review of General Psychology*, 10892680211024124. https://doi.org/10.1177/10892680211024123

Resnick, B. (2017, November 30). *All Muslims are often blamed for single acts of terror. Psychology explains how to stop it.* Vox. https://www.vox.com/science-and-health/2017/11/30/16645024/collective-blame-psychology-muslim

Sacchi, S., Rusconi, P., Russo, S., Bettiga, R., & Cherubini, P. (2012). New knowledge for old credences: Asymmetric information search about in-group and out-group members. *British Journal of Social Psychology*, *51*(4), 606–625. https://doi.org/10.1111/j.2044-8309.2011.02026.x

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The Parasocial Contact Hypothesis. *Communication Monographs*, *72*(1), 92–115. https://doi.org/10.1080/0363775052000342544

Shaver, J. H., Sibley, C. G., Osborne, D., & Bulbulia, J. (2017). News exposure predicts anti-Muslim prejudice. *Plos One*, *12*(3), e0174606. https://doi.org/10.1371/journal.pone.0174606

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

Smalarz, L., Madon, S., & Turosak, A. (2018). Defendant Stereotypicality Moderates the Effect of Confession Evidence on Judgements of Guilt. *Law and Human Behavior*, *42*(4), 355–368. https://doi.org/10.1037/lhb0000286

Stephan, W. G., Boniecki, K. A., Ybarra, O., Bettencourt, A., Ervin, K. S., Jackson, L. A., McNatt, P. S., & Renfro, C. L. (2002). The role of threats in the racial attitudes of Blacks and White. *Personality and Social Psychology Bulletin*, *28*(9), 1242–1254. https://doi.org/10.1177/01461672022812009

Stephan, W. G., & Stephan, C. W. (1996). Predicting prejudice. *International Journal of Intercultural Relations*, *20*(3–4), 409–426. https://doi.org/10.1016/0147-1767(96)00026-0

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Muslim Immigration, Critical Events, and the Seeds of Majority Members' Support for Radical Responses: An Interactionist Perspective. *Personality and Social Psychology Bulletin*, *45*(1), 133–145. https://doi.org/10.1177/0146167218780989

Taylor, D. M., & Jaggi, V. (1974). Ethnocentrism and Causal Attribution in a South Indian Context. *Journal of Cross-Cultural Psychology*, *5*(2), 162–171. https://doi.org/10.1177/002202217400500202

Trope, Y., & Thompson, E. P. (1997). Looking for truth in all the wrong places? Asymmetric search of individuating information about stereotyped group members. *Journal of Personality and Social Psychology*, *73*(2), 229–241. https://doi.org/10.1037/0022-3514.73.2.229

Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting Portrayals of Race/Ethnicity on Primetime Television over a 20-Year Span and Their Association with National-Level Racial/Ethnic Attitudes. *Communication Faculty Articles and Research*. https://doi.org/10.1111/josi.12094

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), 1409–1413. https://doi.org/10.1016/j.jesp.2008.04.003

Upel, J. W., Jana Brauer, Olaf Biernat, Josefine. (2022, May 13). *Geplanter Anschlag in Essen: 16-Jähriger in U-Haft*. https://www1.wdr.de/nachrichten/ruhrgebiet/polizeieinsatz-schulen-essen-borbeck-100.html

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual violence: '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

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, *73*(1), 73–90. https://doi.org/10.1037/0022-3514.73.1.73

Appendix

Table 1. Frequently named information categories as a function of crime type.

Category	Description & examples	Sexual offence (*n* = 155)	Theft (*n* = 156)
Crime circumstances & motive	Information on the exact circumstances of the crime, crime scene, time of crime, reasons and motives of the perpetrator as well as further influence factors (e.g., alcohol, drugs, weapons, or other involved persons)	55 (35.5%)	88 (56.4%)
Sociodemographics	Information on sociodemographics of the perpetrator: Gender of the perpetrator, place of living, age and name of the perpetrator (even though the latter two were mentioned as part of the instruction)	58 (37.4%)	39 (25.0%)
Ethnicity	Information about the nationality, citizenship, (cultural) origin, ethnicity, religious affiliation, migration background, integration & asylum status of the perpetrator	44 (28.4%)	33 (21.2%)
Occupation & socioeconomic status	Information about the (un)employment and the type of professional activity, socioeconomic status in financial terms, material living situation, social class of the perpetrator	41 (26.5%)	25 (16.0%)
Criminal record	Information on possible previous convictions or whether he was already known to the police before the crime	37 (23.9%)	30 (19.2%)
(Current) social environment	Information about the perpetrator's social contacts and relationships. Not including family relationships.	12 (7.7%)	26 (16.7%)
Perpetrator's physical characteristics	Information about physical characteristics of the perpetrator (e.g., height, appearance, hair or eye colour, clothing, or other aspects of his appearance)	10 (6.5%)	24 (15.4%)
Health status	Information about the perpetrator's health. Includes both physical and psychological (previous) illnesses and noticeable problems	24 (15.5%)	7 (4.5%)

Victimology	Information about the victim, perpetrator – victim relationship: Did they know each other before? Where did they possibly meet before?	19 (12.3%)	2 (1.3%)
Civil status	Information about the perpetrator's relationship / civil status, family / potential children	29 (18.7%)	15 (9.6%)

Note. Participants had the possibility to mention several categories. This table depicts categories which were named at least by 10% of the subsample in one of the two conditions. The numbers indicate the absolute counts. The percentages indicate how many of one condition mentioned at least one association targeting the categories (e.g., 35.5% of the participants in the rape condition generated categories which were clustered into the crime circumstances and motive category (64.5% did not indicate any associations targeting this category) and 56.4% in the theft condition indicated (43.6% did not indicate) associations targeting this category).

Table 2. Descriptive and test statistics for all interest and stereotypicality ratings as a function of perpetrator's group membership and crime type.

Variable	Ingroup perpetrator	Outgroup perpetrator	Theft	Sexual offence	ANOVA				
	M (SD)	*M (SD)*	*M (SD)*	*M(SD)*	Effect	F Ratio	*df*	*p*	η^2_G
Ethnic Background									
Interest	3.22 (1.90)	3.34 (1.94)			p	4.44	1,309	.036	.001
			3.12 (1.92)	3.44 (1.92)	c	2.19	1,309	.140	.007
					p × c	0.05	1,309	.830	.000
Stereotypicality	3.10 (1.79)	5.78 (1.47)			p	480.46	1,309	< .001	.419
			4.47 (2.1)	4.41 (2.06)	c	0.22	1,309	.642	.000
					p × c	0.37	1,309	.541	.001
Religious Affiliation									
Interest	2.83 (1.77)	3.02 (1.93)			p	9.37	1,309	.002	.003
			2.62 (1.80)	3.23 (1.87)	c	9.66	1,309	.002	.027
					p × c	2.44	1,309	.119	.001
Stereotypicality	3.12 (1.73)	5.67 (1.59)			p	392.97	1,309	< .001	.373
			4.28 (2.13)	4.51 (2.06)	c	2.90	1,309	.089	.005
					p × c	2.67	1,309	.103	.004
Cultural Background									

Variable	Ingroup perpetrator	Outgroup perpetrator	Theft	Sexual offence	ANOVA				
	M (SD)	*M (SD)*	*M (SD)*	*M(SD)*	**Effect**	**F Ratio**	***df***	***p***	η^2_G
Interest	3.72 (1.93)	3.83 (1.99)			p	3.50	1,309	.062	.001
			3.59 (1.95)	3.96 (1.95)	c	3.03	1,309	.083	.009
					p × c	0.19	1,309	.664	.000
Stereotypicality	3.52 (1.81)	5.84 (1.39)			p	394.06	1,309	< .001	.342
			4.56 (2.01)	4.8 (1.95)	c	2.82	1,309	.094	.005
					p × c	0.00	1,309	.950	.000
Childhood and Youth									
Interest	5.19 (1.84)	5.07 (1.82)			p	5.80	1,309	.017	.001
			5.02 (1.88)	5.24 (1.77)	c	1.26	1,309	.262	.004
					p × c	1.62	1,309	.204	.000
Stereotypicality	4.77 (1.74)	4.05 (1.66)			p	42.88	1,309	< .001	.043
			4.43 (1.70)	4.39 (1.78)	c	0.08	1,309	.774	.000
					p × c	0.21	1,309	.648	.000
Civil Status									
Interest	4.4 (1.92)	4.32 (1.94)			p	1.28	1,309	.258	.000
			4.16 (1.88)	4.56 (1.96)	c	3.78	1,309	.053	.011
					p × c	4.05	1,309	.045	.001
Stereotypicality	3.98 (1.71)	3.78 (1.57)			p	3.57	1,309	.060	.004
			3.95 (1.63)	3.81 (1.66)	c	0.89	1,309	.347	.002
					p × c	0.00	1,309	.958	.000
Occupation									
Interest	4.75 (1.76)	4.73 (1.77)			p	0.16	1,309	.692	.000
			4.88 (1.67)	4.61 (1.85)	c	1.96	1,309	.162	.006
					p × c	0.00	1,309	.954	.000
Stereotypicality	4.10 (1.76)	3.60 (1.66)			p	23.03	1,309	< .001	.022
			4.15 (1.17)	3.55 (1.70)	c	14.31	1,309	< .001	.031
					p × c	1.63	1,309	.202	.002
Current Social Environment									
Interest	5.32 (1.69)	5.37 (1.62)			p	1.04	1,309	.309	.000

Variable	Ingroup perpetrator	Outgroup perpetrator	Theft	Sexual offence	ANOVA				
	M (SD)	*M (SD)*	*M (SD)*	*M(SD)*	**Effect**	**F Ratio**	***df***	***p***	η^2_G
			5.32 (1.64)	5.37 (1.67)	c	0.09	1,309	.767	.000
					p × c	0.07	1,309	.797	.000
Stereotypicality	4.82 (1.75)	4.76 (1.66)			p	0.44	1,309	.506	.000
			4.82 (1.72)	4.76 (1.69)	c	0.11	1,309	.742	.000
					p × c	1.85	1,309	.175	.001
Health Status									
Interest	4.93 (1.84)	4.85 (1.85)			p	3.62	1,309	.058	.000
			4.87 (1.8)	4.91 (1.89)	c	0.46	1,309	.830	.000
					p × c	0.00	1,309	.995	.000
Stereotypicality	4.06 (1.96)	3.09 (1.61)			p	83.29	1,309	< .001	.068
			3.44 (1.82)	3.72 (1.89)	c	2.66	1,309	.104	.006
					p × c	0.18	1,309	.672	.000
Educational Level									
Interest	4.63 (1.83)	4.59 (1.83)			p	0.65	1,309	.421	.000
			4.64 (1.89)	4.59 (1.77)	c	0.06	1,309	.801	.000
					p × c	0.10	1,309	.759	.000
Stereotypicality	4.82 (1.71)	4.71 (1.64)			p	1.17	1,309	.280	.001
			4.78 (1.71)	4.75 (1.64)	c	0.04	1,309	.852	.000
					p × c	1.32	1,309	.252	.001

Note. *N* = 311. ANOVA = Analysis of Variance; p = perpetrator; c = crime type

Summary of Manuscript 2

Khosrowtaj, Z., Süssenbach, P., & Teige-Mocigemba, S. (2023). The black sheep as a shift of blame? Blaming the ingroup perpetrator and victim in an intergroup deviant setting. *Manuscript Submitted*.

Following the literature on black sheep effect (Marques & Paez, 1994; Marques & Yzerbyt, 1988), we tested whether German participants would indicate harsher judgments towards an ingroup perpetrator than an outgroup perpetrator. We further took into consideration the impact of guilt certainty and whether depending on guilt being (not) certain participants would indicate a black sheep effect (outgroup discrimination). Recall that van Prooijen (2006) observed a black sheep effect when guilt was certain whereas outgroup discrimination occurred when guilt was uncertain, leaving room for the so called benefit of the doubt (van Prooijen, 2006). For guaranteeing the maximal similarity to the original experiment which observed the benefit of the doubt (van Prooijen, 2006) we also implemented a theft condition similar to the one of the original study. The first experiment of this manuscript is the first experimental approach for testing the interacting effect of perpetrator's ethnicity and crime type while taking into account guilt certainty. In Experiment 1, we conducted a 2 (perpetrator ethnicity: German vs. Syrian refugee) × 2 (guilt certainty: certain vs. uncertain) × 2 (crime type: rape vs. theft) between-subjects design. Results indicated a black sheep effect independent of the guilt certainty manipulation. This was apparent only on the affective retribution items used by van Prooijen (2006), that is participants indicated a higher degree of anger and hostility towards the ingroup perpetrator than his outgroup counterpart. We did not observe the interaction between ethnicity and guilt certainty (black sheep effect when guilt is certain and outgroup discrimination when guilt is uncertain) even though the guilt manipulation was successful. This raised the question of another underlying factor for the differential judgments of social groups such as ethnicities.

Given the literature on dehumanization and infrahumanization reviewed as part of the theoretical background of this dissertation, we came up with the idea, that participants may perceive deviant acts as more typical for the outgroup and endorsing an unconditional black sheep effect may be a result of this. As such, if participants view the ingroup as more human in general, a decrease in infrahumanization may vanish the black sheep effect. We tested whether perceiving the outgroup as less human would be the cause of differences in judgments of in- and outgroup perpetrators. As such, as part of Experiment 2, we

manipulated infrahumanization for testing whether a decrease in infrahumanization (i.e., higher perceptions of humanness of the outgroup) may come along with equal or harsher treatments of the outgroup.

We further tested again the effect of guilt certainty for drawing conclusions across both Experiments and thus we conducted a 2 (perpetrator ethnicity: German vs. Afghan refugee) × 2 (guilt certainty: certain vs. uncertain) × 2 (infrahumanization: ingroup helping outgroup vs. outgroup helping ingroup) between subjects' design. In addition, we implemented the question regarding the interest towards the different information categories for testing whether participants would be more interested in receiving more information on stereotypic information categories after working through the newspaper article. We predicted that in line with the results of the first manuscript, participants would indicate higher interest towards the stereotypic information categories (religious affiliation, ethnic background and cultural background) after working through the newspaper article depicting the outgroup (than ingroup) perpetrator. Results of Experiment 2 brought up some unpredicted but interesting patterns regarding infrahumanization. Recall that based on previous literature people tend to ascribe fewer secondary emotions independent of valence to outgroups (Leyens et al., 2000, 2001). Our manipulation of infrahumanization (based on Davies et al., 2018) was supposed to decrease infrahumanization (i.e., leading to higher ascriptions of secondary positive *and* negative emotions to the outgroup). However, we observed a higher ascription of secondary positive emotions to the outgroup following the infrahumanization manipulation, whereas participants ascribed significantly less secondary negative emotions towards the outgroup. One explanation for this pattern may be due to the fact that the infrahumanization manipulation described the outgroup in a prosocial manner and the proceeding negative emotions somehow contradicted the first positive laden newspaper article. Davies et al. (2018) for instance only assessed negative primary and secondary emotions after confronting participants with a natural disaster (negative incident). In sum our results indicate that valence of the emotions may play a crucial role in specific contexts which needs further investigations. Independent of the humanization manipulation, the black sheep effect was observed: participants indicated harsher judgments towards the ingroup perpetrator than the outgroup perpetrator. However, when investigating further correlative patterns, with increased ascription of positive secondary emotions we observed participants being less punitive towards the outgroup perpetrator. We further replicated the victim blaming pattern of Experiment 1, that is, participants indicated higher victim blaming when the perpetrator stemmed from the

outgroup. In addition, the correlative pattern regarding the ascribed positive secondary emotions mirrored the perpetrator judgments. With increasing attribution of secondary positive emotions to the outgroup, the victim received more blame but only when the perpetrator stemmed from the outgroup. And finally, the degree of interest towards different information categories (see also manuscript 1) indicated indeed higher interest towards the perpetrator's religious affiliation, ethnic background and cultural background as well as civil status in case of an outgroup than an ingroup perpetrator. These two Experiments point to a possible shift of blame from the outgroup perpetrator to the ingroup perpetrator (black sheep effect) and victim which may be due to the perceptions of less humanness of the outgroup. This may be in line with the finding that higher ascriptions of positive secondary emotions came along with less severe judgments of the outgroup perpetrator but more severe judgments of the victim (but: correlative pattern).

Manuscript 2

The black sheep as a shift of blame?
Blaming the ingroup perpetrator and victim in an intergroup deviant setting

Zahra Khosrowtaj[1], Philipp Süssenbach[2], & Sarah Teige-Mocigemba[1]
[1]University of Marburg, Gutenbergstr. 18, 35037 Marburg, Germany
[2]Fachhochschule des Mittelstands Bielefeld

This research was supported by a grant from the German Research Foundation (Research Training Group 2271: “Breaking Expectations”, project number: 290878970, project 11).

We wish to thank our student assistants Sophie Gosch, Lea Teutenberg, Marion Ihls, Hannah Mackenrodt, Bettina Salger, Katharina Hildebrandt, Lara Neudert, and Lilly Müller for their marvellous assistance in data collection. We would also like to thank Nicole Grandke for her great contribution in planning and conducting Experiment 1 and Violetta Biermann for her support in data collection for Experiment 2.

Correspondence concerning this article should be addressed to Zahra Khosrowtaj. E-mail: zahra.khosrowtaj@uni-marburg.de.
Disclosure statement: No potential conflict of interest was reported by the authors.

ORCID IDs
Zahra Khosrowtaj: 0000-0001-5327-9656
Philipp Süssenbach: 0000-0001-5643-405X
Sarah Teige-Mocigemba: 0000-0002-7516-1293

Abstract

Crimes by refugees are overrepresented in the media and might raise the expectation that refugees behave more deviantly than the ingroup. Often, refugees' cultural and religious background is used to explain their wrongdoings. In two experiments, we investigated the effect of ethnicity of the perpetrator, guilt certainty, and outgroup infrahumanization on perpetrator and victim blaming when judging a sexual offence (Exp. 1) and property crime (Exp. 2). Participants indicated harsher retributive affect (Exp. 1-2) and blame (Exp. 2) towards the ingroup (black sheep effect). Victim blaming was harsher when the perpetrator stemmed from the outgroup (Exp. 1-2), hinting to a shift of blame, discussed in terms of outgroup derogation. A manipulation of infrahumanization (Exp. 2) did not predict blame attributions, but positive secondary emotions affected blame judgments. The role of valence as part of infrahumanization as well as the association between dehumanization and punishment are discussed.

Keywords: black sheep effect, guilt certainty, benefit of the doubt, infrahumanization, guilt attributions, Muslim refugees

"He was so terrible that he was no longer terrible, only dehumanized."

F. Scott Fitzgerald, Tender is the Night

Refugee depiction in the German media became clearly negative after 2015: reports on violent and sex crimes were overrepresented, whereas, for example, property crimes, the most common crime according to crime statistics, were seldomly represented (Maurer et al., 2021). One of the election posters of the right-wing populist party, "Alternative für Deutschland" depicted a woman with partially bare breasts with an accompanying slogan "German women no fair game, got it? Bindingly demanding integration" (*„Sexistisch, rassistisch, widerlich"*, 2021). It targeted stereotypes against male migrants and refugees and instrumentalized women rights to discriminate against them (Vollmar, 2021). Besides the depiction of refugees and migrants as violent perpetrators (Kearns et al., 2019; Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022), portrayals of the outgroup sometimes involve dehumanization. Subtle dehumanisation – i.e., ascribing less humanness – has been observed in media reports about the conditions of refugees during their flight as well as in refugee camps (Siem et al., 2017). With the present work, we aim to investigate factors affecting guilt attributions towards in- and outgroup perpetrators which were not investigated jointly so far. Specifically, as part of our first experiment, we examined the effect of a) perpetrator ethnicity (ingroup vs. outgroup), b) guilt certainty (certain vs. uncertain) as well as c) crime type (sexual offence vs. property crime) on the attribution of blame towards a perpetrator and a victim. Following up on the results of Experiment 1, we investigated the experimental and correlational effect of infrahumanization as a form of dehumanization (Haslam, 2006; Haslam & Loughnan, 2014) in the context of in- and outgroup property offenders. Before turning to the present research, we will briefly review work on factors that affect the coping with ingroup deviance (Otten & Gordijn, 2014), as our framework relies on the psychological primacy of the ingroup (Yzerbyt et al., 2000).

The black sheep effect: Coping with ingroup deviance

When confronted with ingroup deviance, people might react in two ways: justifying / downplaying the misconduct perpetrated by the ingroup, thus, expressing ingroup bias or more harshly rejecting the misconduct and distancing oneself from the black sheep (Marques & Paez, 1994; Marques & Yzerbyt, 1988; Otten & Gordijn, 2014). This latter pattern might seem contradictory. The black sheep effect, however, serves the maintenance of a positive group image by excluding the unexpected deviant member and getting rid of

the bad exemplars who deviated from the ingroup's norms (Yzerbyt et al., 2000). Whether participants will express a black sheep effect (BSE) is crucially moderated by third factors such as guilt certainty.

Guilt certainty and the benefit of the doubt

One factor affecting the perception of deviance is described as the ambiguity of guilt (Otten & Gordijn, 2014). A fascinating pattern observed by van Prooijen (2006) illustrates one of the necessary conditions for the BSE to occur. When guilt was certain, participants were more angry and hostile towards an ingroup compared to an outgroup wrong-doer, but when guilt was uncertain, participants judged the outgroup wrong-doer more harshly as might be expected from an ingroup bias perspective (van Prooijen, 2006).

Hypocrisy in blame attributions due to crime type?

Crime type seems to affect the blame attributed towards victims and perpetrators (Brems & Wagner, 1994): less blame was attributed to the perpetrator in case of theft than in the case of rape. However, also in the context of sexual assault, perpetrators may be judged relatively leniently. Whereas in one study, participants blamed a victim more harshly and exonerated the perpetrator in case of sexual assault compared to theft (Bieneck & Krahé, 2011), in another recent study victim blaming did not differ regarding different crimes (Reich et al., 2021). So far, the interaction between perpetrator's ethnicity and crime type has not been investigated yet. Taking into account the portrayal of Muslim male migrants in the German media following the sexual assaults in Cologne in 2015/2016 (Stürmer et al., 2019; Wigger et al., 2022), the argument has been put forward that sexual violence is perceived as a racialized and Islamised crime (Wigger, 2019). This might crucially influence the way in- and outgroup perpetrators of sexual violence are perceived.

(Infra)humanization as a moderator for the black sheep effect?

In a recent study, Americans perceived Muslims and Arabs as the least evolved human group followed by Mexican immigrants, South Koreans, and Chinese people, while Europeans, Swiss, Japanese and Australians were perceived as similarly evolved (Kteily et al., 2015). Blatant dehumanization and infrahumanization, a more subtle form of dehumanization (Leyens et al., 2000), predicted support for reducing Arab immigration (Kteily et al., 2015).

Following infrahumanization theory, human essence (i.e., language, cognition, and specific emotions) differentiates humans from animals (Leyens et al., 2000, 2001) and people tend to reserve this human uniqueness for their ingroup, thereby "infrahumanizing" outgroups. Thus, independent of their valence, positive secondary emotions (such as hope)

as well as negative ones (such as melancholy) are ascribed to one's ingroup whereas primary emotions that are also experienced by animals (such as fear) are attributed to the outgroup and, thus, infrahumanization remains distinct from mere ingroup favouritism (Haslam, 2006; Leyens et al., 2000, 2001). Perceiving one's ingroup to be collectively responsible for an outgroup mass killing led participants to infrahumanize the outgroup victims (Castano & Giner-Sorolla, 2006). In the context of sexual violence, dehumanizing sexual offenders predicted reduced support for their rehabilitation, higher sentence lengths, and the social exclusion of the offender (Viki et al., 2012).

The present work

With the present work, we shed light on the four above mentioned factors affecting the evaluation of perpetrators and victims. In the first experiment, we examined the effect of perpetrator ethnicity[8], guilt certainty, and crime type on blame ratings. Here, we expected German participants to blame the ingroup (German) perpetrator more harshly compared to the outgroup (Syrian) perpetrator in line with the BSE, when guilt is certain (van Prooijen, 2006). We further expected participants to blame the outgroup perpetrator more harshly when guilt was not certain, thus observing a "benefit of the doubt"-effect. Moreover, we examined whether crime type has an influence on this pattern. Finally, we investigated whether harsher victim blaming and more lenient judgments towards the perpetrator occur in case of rape as observed by Bieneck and Krahé (2011).

In Experiment 2, we investigated the role of infrahumanization on blame judgments in the case of theft. More concretely, we examined whether increasing the humanization of the outgroup might be associated with equal treatment of the perpetrators. Thus, we tested infrahumanization as a possible moderator for the BSE to (not) occur in the context of intergroup wrong doings.

[8] There is a racialization of Islam (Shooman, 2012) even though the ethnicity should not be equalled with one's religious affiliation – "not all Arabs are Muslims nor all Muslims are Arabs" (Kteily et al., 2015, p. 42). Antimuslim racism is prevalent in Germany (Zick, 2017) and Shooman (2012) hints to the synonymous use of categories such as Muslim, Turk, Arab, and Migrant. Thus, in the following, we will refer to perpetrator's ethnicity, while this factor was manipulated using a predominantly Muslim country (Syrian and Afghan) as ethnicity for the outgroup perpetrator.

Experiment 1

Methods

All measures and manipulations relevant for the scope of the present work as well as all exclusions are reported for both experiments.

Sample size estimation

Building on the study by van Prooijen (2006) who observed an interaction effect between group membership and guilt certainty of f = .27 (estimated upon the reported test statistics using a freely available tool by Lakens (2013)), an a-priori power analysis was run. Due to differences to the original study (e.g., van Prooijen operationalized ingroup vs. outgroup via Dutch vs. German soccer teams, fictious companies and membership in differing Dutch universities), we decided for a more conservative effect size and estimated f at.17. Assuming an effect size of f = .17 for the interaction between group × guilt certainty, we required a sample size of 366 participants for a power of 90%. We aimed at collecting data from N = 400 anticipating possible exclusions.

Participants

A total of N = 478 participants completed the online experiment. Due to our a priori determined exclusion criteria, 45 participants had to be excluded: Five participants who were younger than 18 years were excluded (we did not exclude participants who did not indicate their age, n = 5). Moreover, we only analysed data of German participants. Therefore, we excluded 17 participants who indicated having a non-German first language and ten further participants who indicated that at least one parent had a first language from a predominantly Muslim country (i.e., Arabic, Afghan, Farsi). Finally, two participants who had participated in previous similar studies and eleven participants who were reported as extremely fast completers by the survey software (Leiner, 2016) were excluded. This resulted in a total analysis sample of N = 433 participants (n = 279 female, n = 133 male, n = 6 diverse, n = 15 not answered) with an average age of M = 28.47 years and SD = 9.76 (range = 18-75 years). We conducted a sensitivity analysis using this final sample size. For a power of .90, the interaction effect would need to be f = .156 or greater.

Procedure and Measures

Overall procedure. We conducted a 2 (perpetrator ethnicity: German vs. Syrian refugee) × 2 (guilt certainty: certain vs. uncertain) × 2 (crime type: rape vs. theft) between-subjects design. We invited participants to take part in an online study. We randomly assigned participants to one of eight experimental conditions in which they read an alleged newspaper article describing either a rape case or a theft case. Following the article, we

asked participants to indicate how much they blamed the perpetrator and the victim and assessed hostile affect and pity felt towards the perpetrator and victim. For aims beyond the scope of the present work participants additionally worked through some additional measures; in the following we will focus on the measures relevant for the current experiment. A single item checked the guilt certainty manipulation using a 7-point Likert scale ("Finally, a question about the newspaper article you read. Please think back to the content of the article you read. To what extent has [perpetrator's name] guilt been proven in relation to the [crime done]?"), with 1 = not at all and 7 = very strongly. Finally, participants completed demographical variables such as gender, age, first language, first language of their parents, contact to Muslims, educational level, approximate size of their living place, and some questions targeting at whether participants had been a crime victim themselves (e.g., sexual assaults, theft, physical violence, social exclusion). We gave participants the possibility to take part in a lottery for gift cards, thanked them for participation and ended the experiment with a debriefing about the fictional character of the articles.

Newspaper article. We created eight parallel newspaper articles, which only differed regarding the experimental factors. Based on work conducted by van Prooijen (2006, Exp. 3 & Exp. 4) we adapted the bicycle theft scenario and manipulated guilt certainty by adding the information that surveillance cameras recorded the crime and the perpetrator was (vs. was not) detected certainly (van Prooijen, 2006, Exp. 4). The articles depicting rape as crime were constructed following the bicycle articles with the aim of achieving maximum parallelism as possible. One exemplary article (rape, outgroup perpetrator, and guilt certain condition) is presented below.

Hamburg

Rape: Syrian refugee arrested

20.01.2019, 11:00

In the night from December 14, 2018, to December 15, 2018, a rape occurred in a side street near the main train station. The 24-years-old Ines M. from Hamburg, who was the victim of the rape, reported this to the police the following day. Shortly thereafter, investigation procedures were initiated against the 28-years-old Syrian refugee Mustafa K.

Surveillance cameras recorded the course of events. When evaluating the recordings, Mustafa K. could be clearly identified as part of the circumstances of the crime.

Judgments. We assessed participants' judgments regarding the case using 20 items based on previous research, e.g. "How much influence did [he/she] have on the outcome of the situation?", "How likely is the accused guilty of a crime?" (Bieneck & Krahé, 2011; Süssenbach et al., 2012; Süssenbach et al., 2017) and retributive affect, e.g. "How hostile do you feel towards the suspect?", "How angry do you feel toward the suspect?" (van Prooijen, 2006; van Prooijen & Lam, 2007). All dependent measures were assessed on 7-point Likert scales with higher scores indicating higher negative judgments / emotions except for one categorical single item assessing the main guilt: "Who bears the primary responsibility for things happening the way they happened?" with a dichotomous answer depicting the names of the perpetrator and victim.

Results

Guilt certainty manipulation check and preparatory analyses

An independent samples *t*-test with guilt certainty as independent variable and our manipulation check item as dependent variable revealed the expected effect of guilt certainty: Participants in the guilt certain condition indicated higher guilt certainty (M = 5.58, SD = 1.76) than participants in the guilt uncertain condition (M = 1.98, SD = 1.42), $t(415.213) = 23.51$, $p < .001$, $d = 2.26$.

We conducted several factor analyses with all dependent variables (except for the single dichotomous item). Using a PFA procedure and allowing for correlated factors, we determined that a two-factor solution was both theoretically most plausible as well as the best fit to the data. The first (second) factor included items targeting blame and retributive affect towards the perpetrator (victim), Cronbach's Alpha = .90 (.81), 11 items (8 items).

Blame and retributive affect towards the perpetrator and the victim

We ran an ANOVA with perpetrator blame as the dependent variable and perpetrator ethnicity, guilt certainty, and crime type as independent variables. We observed a main effect of guilt certainty, $F(1, 425) = 237.27$, $p < .001$, $\eta_p^2 = .358$, such that participants in the guilt certain condition indicated harsher retributive affect towards the perpetrator (M = 5.31, SD = .98) compared to the participants in the guilt uncertain condition (M = 3.94, SD = 1.23). In addition, we observed a main effect of crime type, $F(1, 425) = 169.69$, $p < .001$, $\eta_p^2 = .285$, such that participants judged the perpetrator more harshly in the rape condition (M = 5.21, SD = 1.23) compared to the theft condition (M = 4.06, SD = 1.11). There was no effect of perpetrator ethnicity, $F(1, 425) = 3.01$, $p = .084$, $\eta_p^2 = .007$.

Contrary to the results observed by van Prooijen (2006), we did not observe any interaction between perpetrator ethnicity × guilt certainty, $p = .961$. Further, we did not observe an interaction between perpetrator ethnicity × crime type, $p = .775$, nor any other interaction effect, guilt certainty × crime type, $p = .711$, perpetrator ethnicity × guilt certainty × crime type, $p = .883$.

To take a closer look, we conducted an additional analysis only with the retributive affect items that were used by van Prooijen (2006) in his Experiments 1-2, (e.g., "How angry [hostile] do you feel toward the perpetrator?). This analysis revealed a main effect of perpetrator ethnicity, $F(1, 425) = 7.77$, $p = .006$, $n_p^2 = .018$, such that participants expressed higher levels of retributive affect towards the ingroup perpetrator ($M = 3.75$, $SD = 1.98$) compared to the outgroup perpetrator ($M = 3.31$, $SD = 1.95$), suggestive of a BSE. Again, we observed a main effect of guilt certainty, $F(1, 425) = 79.58$, $p < .001$, $n_p^2 = .158$, such that participants attributed higher rates of retributive affect in the guilt certain condition ($M = 4.22$, $SD = 1.85$) compared to the guilt uncertain condition ($M = 2.83$, $SD = 1.85$). Also, we observed a main effect of crime type, $F(1, 425) = 101.72$, $p < .001$, $n_p^2 = .193$, such that participants attributed higher retributive affect in the rape condition ($M = 4.32$, $SD = 2.01$) compared to the theft condition ($M = 2.76$, $SD = 1.60$). There was, however, no interaction effect, perpetrator ethnicity × guilt certainty, $p = .872$, perpetrator ethnicity × crime type, $p = .591$, crime type × guilt certainty, $p = .085$, perpetrator ethnicity × guilt certainty × crime type, $p = .525$.

We conducted an ANOVA with victim blame as the dependent variable and the same factors as before in the model. This analysis revealed a significant effect of perpetrator ethnicity, $F(1, 425) = 8.58$, $p = .004$, $n_p^2 = .020$, such that the victim was judged more negatively when the perpetrator stemmed from the outgroup ($M = 2.26$, $SD = 1.02$) than from the ingroup ($M = 2.05$, $SD = .85$). In addition, there was a significant effect of crime type, $F(1, 425) = 115.45$, $p < .001$,$n_p^2 = .214$, such that the victim was judged more negatively in the theft condition ($M = 2.57$, $SD = .95$) than in the rape condition ($M = 1.73$, $SD = .71$). Furthermore, we observed a significant interaction between perpetrator ethnicity and crime type, $F(1, 425) = 5.14$, $p = .024$,$n_p^2 = .012$.

For probing the two-way interaction between perpetrator ethnicity × crime type, we ran separate independent *t*-tests for the theft and rape condition. Participants indicated harsher judgments towards the victim when the perpetrator stemmed from the outgroup ($M = 2.85$, $SD = 1.03$) compared to the ingroup ($M = 2.39$, $SD = .90$) in the theft condition,

$t(237) = -3.62$, $p < .001$. There was no difference between the victim blame judgments in the rape conditions, $t(237) = -.89$, $p = .370$.

No other main nor interaction effects were observed: guilt certainty, $p = .115$, guilt certainty × perpetrator ethnicity, $p < .155$, guilt certainty × crime type, $p = .143$ and perpetrator ethnicity × guilt certainty × crime type, $p = .715$.

Main blame: forced choice judgement

We conducted a binomial logistic regression for estimating the effect of perpetrator's group membership, guilt certainty, and crime type on the likelihood that participants would blame the perpetrator (coded with 1) compared to the victim (coded with 0). The logistic regression model was statistically significant, x^2 (3) = 136.24, $p < .001$ and explained 43.5% (Nagelkerke's R^2) of the variance in main guilt and correctly classified 85.2% of cases. The area under the ROC curve was .863, 95% CI [.819, .906], which is an excellent level of discrimination according to Hosmer et al. (2013). All predictors in the model were statistically significant as shown in Table 1. Participants in the ingroup condition had 1.98 times higher odds to blame the perpetrator as compared to the victim than participants in the outgroup condition. Participants in the guilt certain condition had 18.06 times higher odds to blame the perpetrator as compared to the victim than participants in the guilt uncertain condition. Finally, participants in the theft condition had 11.55 times higher odds to blame the victim than the perpetrator as compared to the rape condition.

Table 1. ***Logistic regression predicting likelihood of blaming the perpetrator based on perpetrator's ethnicity, guilt certainty, and crime type.***

	B	SE	Wald	*df*	*p*	OR	95% CI Odds Ratio
Group	.68	.31	4.94	1	.026	1.98	[1.08, 3.61]
Guilt certainty	2.89	.39	53.65	1	<.001	18.06	[8.33, 39.17]
Crime type	2.45	.35	47.78	1	<.001	11.55	[5.77, 23.11]
Constant	-5.36	.53	101.38	1	<.001	.005	

Note. Categorical codings for the predictors include Perpetrator ethnicity: Ingroup = 1, Outgroup = 0; Guilt certainty: 100% = 1, 50% = 0; crime type: theft = 1, rape = 0.

Discussion

In the present experiment, we investigated the effect of a perpetrator's ethnicity, guilt certainty, and crime type on the attributions of guilt towards the perpetrator and victim. We

observed participants being more punitive towards perpetrators of rape and the manipulation of guilt certainty proved successful. When looking at the retributive affect items used by van Prooijen (2006) our data indicated a BSE (Marques & Paez, 1994; Marques & Yzerbyt, 1988) – in contrast to van Prooijen, however, we did not observe an interaction between guilt certainty and perpetrator ethnicity. The attribution of main guilt was in line with the BSE as the odds of blaming the perpetrator compared to the victim were higher in the ingroup than in the outgroup perpetrator condition.

In general accordance with a BSE, participants blamed the victim less when the perpetrator stemmed from the ingroup than from the outgroup. In addition, we observed participants indicating higher rates of victim blaming in the theft than in the rape condition, which was not predicted based on results observed by Bieneck and Krahé (2011). This might be in support with the observation that participants were less likely to blame rape victims than theft victims when blame is assessed directly compared to indirectly (Felson & Palmore, 2018), both the direct and indirect measure involved explicit measures but differed in their wording.

This first experiment did not confirm our main hypothesis, namely that guilt certainty moderates the occurrence of the BSE (Otten & Gordijn, 2014; van Prooijen, 2006). Different from van Prooijen (2006), we only observed a BSE as a main effect on retributive affect independent of guilt certainty. Despite also employing theft as a crime as implemented by van Prooijen (2006), guilt certainty did not affect the occurrence of a positive or negative bias as reviewed by Otten and Gordijn (2014). The question arises whether there might be another factor that might affect the nature of bias in case of social outgroups such as ethnic and national minorities. More precisely, it could be reasoned that participants exhibited an unconditional BSE in the first experiment not only because they wanted to distance themselves from the ingroup harm doer but also because they believed that the conducted harm is more typical for the outgroup as a whole (Khosrowtaj, Teige-Mocigemba, et al., 2023; Khosrowtaj, Yzerbyt, et al., 2023).

(Infra)humanization of the outgroup and the evaluation of deviant members

If participants of our first experiment indicated harsher judgments towards the ingroup perpetrator because they viewed the ingroup as more human in general, whereas the outgroup, and thus, the outgroup perpetrator was infrahumanized, then reducing the infrahumanization of the outgroup might change the subsequent guilt attributions towards the outgroup perpetrator. In consequence, the BSE might vanish as both groups are viewed as similarly human and moral following the manipulation.

Bastian and colleagues (2011) demonstrated that dehumanization might in fact be associated with less punishment (but see Bastian et al., 2013). Perceiving perpetrators as unable to act morally leads to their exoneration as they lack human uniqueness traits (Bastian et al., 2011). In Experiment 2, we investigated the possibility that the ingroup deviant is perceived as morally capable and thus the BSE occurred while the outgroup deviant might be perceived as morally not capable thus not deserving punishment which would be in line with results discussed by Bastian and colleagues (2011) and the dehumanization of Muslims as part of the media and in European and US samples (Bruneau et al., 2018; Kteily et al., 2015; Wigger, 2019; Wigger et al., 2022). To test this reasoning, we experimentally manipulated infrahumanization in the theft crime condition in our second experiment. We relied on previous work conducted by Davies et al. (2018) which found that an outgroup helping the ingroup after a natural disaster decreased infrahumanization.

To sum up, in our second experiment, we tested whether the BSE would vanish following the decrease of outgroup infrahumanization. In other words, we tested the hypothesis that reducing infrahumanization might lead to *equal* treatment of in- and outgroup perpetrators or even harsher treatment of the outgroup perpetrator. Thus, we predicted that in the control condition, the ingroup relative to the outgroup perpetrator is judged more harshly (i.e., BSE). This effect should be absent in the experimental condition in which the outgroup is humanized. In case that the experimental manipulation is not successful, we further tested this idea by examining correlative patterns. If infrahumanization is crucial for the emergence of a black sheep effect, secondary emotions ascribed to the outgroup should be positively correlated with the blaming of an outgroup perpetrator (i.e., humanization leading to less lenient judgments). In addition, we introduced an explorative measure of information search to further examine the hypothesis that participants are more interested in stereotypic information categories (e.g., religious affiliation) in case of the outgroup than the ingroup perpetrator. We planned to explore whether participants endorse higher interest in receiving further information about stereotypic information categories such as religious affiliation, ethnic background, and cultural background in the outgroup condition than in the ingroup condition.

Experiment 2

Methods

Sample size estimation

In Experiment 1, the BSE for the retributive affect items was n_p^2 = .018, i.e., f = .135. However, as we investigated the effect of a new experimental manipulation on guilt attributions, we decided for a more conservative effect size estimate. Thus, assuming a small interaction between infrahumanization × perpetrator ethnicity of f = .10 we conducted an a-priori sample size estimation. This analysis revealed a total *N* of 1053. We aimed at collecting *N* = 1100 participants for compensating possible exclusions.

Participants

A total of *N* = 1159 participants completed the online experiment. Exclusion criteria were as in Experiment 1. We excluded ten participants who were younger than 18 years including those with suspicious entries like "-1" or "-15" but not one participant with an entry as "333" which we assumed to stand for 33 as the other demographical entries seemed to be unobtrusive (we did not exclude three participants that did not indicate their age). We further excluded 37 participants who indicated a non-German first language. In addition, we excluded 22 participants who endorsed that at least one of their parents had a mother tongue from a predominantly Muslim country (i.e., Arabic, Afghan, Farsi), and nine participants who had participated in previous similar studies. Finally, we excluded 17 participants who were detected as extremely fast completers by the survey software (Leiner, 2016). We further excluded participants based on three attention checks regarding the first newspaper article (for further details, see below). Here, we excluded participants who selected a wrong answer in all the three questions (*n* = 1) and those indicating not remembering the information in all the three questions (*n* = 19). With this, we aimed at excluding participants who did not read the first article (infrahumanization). This resulted in a total analysis sample of *N* = 1044 participants (*n* = 732 female, *n* = 279 male, *n* = 10 diverse, *n* = 23 not answered). Only for the calculation of the average age the participant who indicated 333 was excluded, this resulted in an average age of *M* = 32.02 years and *SD* = 15.79 (range = 18-80 years). In addition, we conducted a sensitivity analysis with this final sample size. With our analysis sample *N* = 1044, error rate = 5% and a power of 90% we were able to detect effect sizes as little as f = .10.

Procedure and Measures

Overall procedure. We conducted a 2 (perpetrator ethnicity: German vs. Afghan refugee) × 2 (guilt certainty: certain vs. uncertain) × 2 (infrahumanization: ingroup helping

outgroup vs. outgroup helping ingroup) between subjects' design. We asked participants to read two newspaper articles and to answer questions related to both articles. Participants first read a newspaper article about helping behaviour during the COVID-19 pandemic. This article contained a manipulation aimed at influencing infrahumanization. Next, participants completed text comprehension items as well as a measure of infrahumanization. Subsequently, participants read a second newspaper article that described a theft. This article was the same as in Experiment 1. Following the newspaper article, participants responded to dependent variables on blame attributions to the perpetrator and the victim. Finally, participants responded to two manipulation check items on guilt certainty, filled in demographics and were given the choice to participate in a gift card lottery before they were thanked and fully debriefed.

Manipulation of infrahumanization. Following previous work (Davies et al., 2018), participants were randomly assigned to one of two conditions. Participants read that the first newspaper article would deal with behaviour during the COVID-19 pandemic. Then, participants either read a newspaper article inspired and adapted from an existing German online newspaper article (Förster, 2020) describing an Afghan family sewing 360 face masks for a clinic and a seniors' citizen home (outgroup helping the ingroup condition) or about a German family sewing 360 masks for a refugee accommodation (ingroup helping the outgroup). Both conditions were parallel except for the names and, accordingly, the nationality of the helping party and the help receiving organisation.

Attention check and infrahumanization measure. After reading the article, we asked participants three multiple choice attention check items regarding the article (e.g., "How many masks did the couple sew?") with twofold aims: 1) we intended to exclude subjects who failed to answer the attention items correctly and 2) having distractor items to separate the manipulation from the infrahumanization measure. For the infrahumanization measure, participants were told that scientific results indicate that different groups of people are subscribed different emotions and that there are research gaps for some groups. We asked participants to rate in how far the group of Afghans experience specific emotions on a 6-point Likert scale, from 1 = not at all to 7 = very strongly. We presented three primary and three secondary positive (e.g., joy, hope) and negative emotions (e.g., pain, melancholy), respectively, used by Rohmann et al. (2009); that is, participants completed 12 items in total that were presented in a fixed alternating order due to valence and uniqueness order.

Dependent variables. We used a subset of the dependent measures that were used in Experiment 1. Based on an exploratory factor analysis and reliability analysis we computed

three indices: perpetrator blaming (Cronbach's α = .91, five items), victim blaming (Cronbachs α = .81, three items), and retributive judgement towards the perpetrator (Cronbach's α = .88, five items).

Information seeking. In the information seeking measure, participants viewed nine information categories about the perpetrator (e.g., religious affiliation, educational level) and rated their interest in having more information on the specific categories on a 7-point Likert scale ranging from 1 = "does not interest me at all" to 7 = "interests me very strongly". This measure was tested as part of another online experiment (Khosrowtaj, et al., 2022) as a subtle way of stereotypic information seeking.

Manipulation check for guilt certainty. As manipulation check for guilt certainty, we asked participants two questions: "Please think back to the content of the last article you read. To what extent has the suspect's guilt in relation to the thefts been proven?"; "How convinced are you personally of the reported guilt of the suspect presented as part of the newspaper article which you read later?". Both items used a slider ranging from 0% to 100% with higher values representing higher guilt certainty.

Results

Effect of infrahumanization manipulation on attribution of emotions

We conducted independent samples *t*-tests to test the effect of the infrahumanization manipulation on the attribution of emotions. Regarding primary and secondary emotions regardless of valence, we observed no effects of the manipulations, both *ps* > .861.

We observed no effect of the manipulation on positive nor negative primary emotions, both *p*s > .32. The manipulation, however, influenced the attribution of secondary positive emotions as intended, $t(1042) = 2.71$, $p = .007$, $d = .17$. Participants attributed more secondary positive emotions to Afghans in the outgroup helping condition ($M = 5.19$, $SD = 1.21$) than in the ingroup helping condition ($M = 4.99$, $SD = 1.26$). Not in line with prior research, participants attributed fewer secondary negative emotions to Afghans in the outgroup helping condition that intended to humanize the outgroup ($M = 4.38$, $SD = 1.60$) compared to the helping ingroup condition ($M = 4.58$, $SD = 1.46$), $t(1042) = -2.05$, $p = .041$, $d = -.13$. An overall explorative examination of the ascription of emotions independent of valence and uniqueness (personal communication Michal Bilewicz, August 2022) was not significant, $p = .717$.

Manipulation check for guilt certainty

Two *t*-tests confirmed that the guilt certainty manipulation worked. Participants judged the article to contain more evidence proving the suspect's guilt in the guilt certain condition (M = 82.94 %, SD = 27.54 %) than in the guilt uncertain condition (M = 14.27 %, SD = 20.37 %), $t(957{,}79) = 45.79$, $p < .001$, $d = 2.84$. In addition, participants were subjectively more convinced of the suspect's guilt in the guilt certain condition (M = 79.62 %, SD = 26.23 %) than in the guilt uncertain condition (M = 19.43 %, SD = 24.06 %), $t(1042) = 38.65$, $p < .001$, $d = 2.39$.

Effects of infrahumanization manipulation, guilt certainty, and perpetrator ethnicity on guilt attributions to the perpetrator and victim

We conducted ANOVAs with infrahumanization, guilt certainty, and perpetrator ethnicity as independent variables predicting blame attributed to the perpetrator (retributive judgement). As in Experiment 1, we observed a significant main effect of guilt certainty, $F(1, 1036) = 1230.11$, $p < .001$, $\eta_p^2 = .543$ ($F(1, 1036) = 253.33$, $p < .001$, $\eta_p^2 = .196$). Participants indicated harsher blame ratings in the guilt certain condition, $M = 5.71$, $SD = 1.02$ ($M_{retributive\ judgement} = 3.57$, $SD_{retributive\ judgement} = 1.24$) compared to the guilt uncertain condition, $M = 3.24$, $SD = 1.26$ ($M_{retributive\ judgement} = 2.39$, $SD_{retributive\ judgement} = 1.21$). We further observed a significant main effect of perpetrator ethnicity, $F(1, 1036) = 22.05$, $p < .001$, $\eta_p^2 = .021$ ($F(1, 1036) = 43.19$, $p < .001$, $\eta_p^2 = .040$). Participants indicated harsher blame ratings towards the ingroup perpetrator, $M = 4.64$, $SD = 1.61$ ($M_{retributive\ judgement} = 3.22$, $SD_{retributive\ judgement} = 1.34$) compared to the outgroup perpetrator, $M = 4.31$, $SD = 1.75$ ($M_{retributive\ judgement} = 2.73$, $SD_{retributive\ judgement} = 1.33$), revealing a BSE. Unsurprisingly, the failed infrahumanization manipulation had neither a main effect, $F(1, 1036) = .05$, $p = .830$, nor did it interact with any of the other experimental factors, all ps > .098.

The interaction between perpetrator ethnicity and guilt certainty was significant for retributive judgement, $F(1, 1036) = 6.65$, $p = .010$, $\eta_p^2 = .006$. For probing this interaction, we ran two independent sample *t*-tests with perpetrator ethnicity as the independent factor and looking at the data separately at each level of guilt certainty. Both simple effects indicated a BSE. However, the BSE was stronger for the guilt certain condition, $t(519) = 6.43$, $p < .001$, $M_{ingroup} = 3.90$, $SD_{ingroup} = 1.21$, $M_{outgroup} = 3.23$, $SD_{outgroup} = 1.18$, than for the guilt uncertain condition, $t(521) = 2.79$, $p = .005$, $M_{ingroup} = 2.53$, $SD_{ingroup} = 1.11$, $M_{outgroup} = 2.24$, $SD_{outgroup} = 1.29$. We did not observe any effect of the infrahumanization manipulation nor an interaction of this manipulation with one of the other factors for the retributive judgments, all ps > .091.

To evaluate possible effects on victim blaming, we ran an ANOVA for victim blaming and the same factors as in the previous models. This analysis revealed no effect of guilt certainty, F(1, 1036) = 3.20, p = .074, and no effect of infrahumanization, F(1, 1036) = 2.25, p = .134. Replicating Experiment 1, we observed a significant effect of perpetrator ethnicity, F(1, 1036) = 5.12, p = .024, $\boldsymbol{\eta_p^2}$ = .005, such that participants blamed the victim more harshly when the perpetrator stemmed from the outgroup (M = 3.06, SD = 1.45) than the ingroup (M = 2.94, SD = 1.48). We did not observe any interaction effect, all ps > .28.

Effect of ascribed secondary emotions on guilt attributions

As the experimental manipulation of infrahumanization was not successful and did not affect guilt attributions, we conducted multiple regression analyses to explore whether ascribing more versus less secondary emotions to the outgroup has downstream consequences for the assignment of blame towards an individual perpetrator and victim.[9] Here, we created contrasts for perpetrator ethnicity (ingroup = -0.5 and outgroup = 0.5), guilt certainty (certain = -0.5, uncertain = 0.5) and centered the secondary positive and negative emotions which we used as predictors. We conducted a multiple regression analysis for predicting perpetrator blaming (retributive judgement) using the experimental factors and the centered emotions as predictors. Besides the aforementioned significant main effects of perpetrator ethnicity and guilt certainty, we further observed a significant interaction between secondary positive emotions and perpetrator ethnicity on perpetrator blaming (retributive judgement), b = -.15, *t*(1033) = -2.19, *p* = .029 (b = -.23, *t*(1033) = -3.31, *p* < .001). Simple slope analyses revealed a significant effect of the secondary positive emotions for the outgroup perpetrator: b = -.11, *t*(1033) = -2.32, *p* = .021 (b = -.15, *t*(1033) = -2.89, *p* = .004) but not for the ingroup perpetrator: b = -.03, *t*(1033) = .73, *p* = .463 (b = .08, *t*(1033) = 1.75, *p* = .081). Thus, with increasing attribution of secondary positive emotions, participants blamed the outgroup perpetrator less severely.

Mirroring these patterns, another regression analysis for victim blaming indicated a significant interaction between perpetrator × secondary positive emotions, *t*(1033) = 3.23, *p* = .001. Simple slope analysis revealed no effect of secondary positive emotions on victim blaming when the perpetrator stemmed from the ingroup, b = -.07, *t*(1033) = -1.25, *p* =

[9] Two valence-independent analyses indicated no significant interaction between the perpetrator × primary emotions, *p* > .207 nor the perpetrator × secondary emotions, *p* > .720. Thus, we continued with the following analyses including the secondary positive and negative emotions where we also observed differential patterns due to the experimental aid conditions.

.212, but there was a significant effect of secondary emotions on victim blaming when the perpetrator stemmed from the outgroup, b = .19, $t(1033) = 3.26$, $p = .001$. Thus, with increasing attribution of secondary positive emotions to the outgroup (Afghans), participants blamed the victim more when the perpetrator stemmed from the outgroup.

Effect of perpetrator's group membership on information seeking behaviour

We ran nine independent samples *t*-tests for the nine information categories presented to test the impact of perpetrator ethnicity on information seeking. Interestingly, we observed participants indicating higher interest towards the categories ethnic background, religious affiliation, cultural background, and civil status in case of an outgroup perpetrator compared to an ingroup perpetrator. Table 2 depicts these results.

Table 2. ***Information seeking behaviour as a function of perpetrator group membership.***

Information category	*M* (*SD*) ($n_{ingroup}$ = 527)	*M* (*SD*) ($n_{outgroup}$ = 517)	*t*(*df*)	Two-sided *p*
Educational level	3.61 (2.09)	3.63 (2.17)	-.19 (1042)	= .848
Childhood & youth	4.02 (2.19)	3.91 (2.23)	.83 (1042)	= .406
Current social environment	4.70 (2.07)	4.72 (2.11)	-.16 (1042)	= .870
Ethnical background	2.03 (1.63)	2.32 (1.77)	-2.68 (1031.97)	= .007
Religious affiliation	1.57 (1.24)	1.86 (1.49)	-3.33 (1001.17)	< .001
Civil status	2.60 (1.85)	3.07 (2.11)	-3.79 (1020.08)	< .001
Health status	3.40 (2.09)	3.42 (2.19)	-.16 (1036.98)	= .873
Occupation	3.45 (2.02)	3.57 (2.19)	-.93 (1031.76)	= .355
Cultural background	2.29 (1.71)	2.78 (1.95)	-4.35 (1018.37)	< .001

Discussion

In Experiment 2, we investigated whether manipulating infrahumanization via intergroup helping behaviour might diminish the BSE, in other words, whether ascribing more secondary emotions to the outgroup would lead to equal judgment of an outgroup perpetrator. Our experimental manipulation did not work as expected. Participants in the outgroup helping the ingroup condition ascribed slightly more secondary positive emotions to the outgroup compared to the ingroup, but at the same time ascribed fewer secondary negative emotions. Based on previous literature reviewed above, secondary emotions are attributed independent of valence more often to the ingroup than to the outgroup (Leyens et al., 2000, 2001). Thus, we expected the humanization condition to increase both secondary positive *and* negative emotions.

One explanation for this unexpected pattern may be that the negative secondary emotions assessed did somehow contradict the positive laden first newspaper article in which the outgroup helped the ingroup. However, one could argue that this should affect primary negative emotions in a similar fashion, which was not the case. However, Davies et al., (2018) observed a reduction of infrahumanization following an outgroup helping the ingroup after a natural disaster while assessing only negative (primary and secondary) emotions. We will elaborate on the role of valence in research on infrahumanization within the General Discussion.

Independent of the humanization manipulation, the BSE was observed: Participants indicated harsher judgments (blame and retributive judgement) towards the ingroup perpetrator than the outgroup perpetrator. In addition, the ascription of secondary positive emotions affected the judgments towards the perpetrator (correlational): With increased attribution of positive secondary emotions, participants were less punitive towards the outgroup perpetrator.

Regarding victim blaming, participants blamed the victim more when the perpetrator stemmed from the outgroup, replicating the pattern observed in Experiment 1. Further, we observed an interaction with secondary positive emotions, such that with increasing attribution of secondary positive emotions to the outgroup, the victim received more blame – but only if the perpetrator stemmed from an outgroup.

Our explorative investigation of information search revealed a bias: participants were more interested to learn about the perpetrator's religious affiliation, ethnic background, and cultural background (as well as the civil status) when the perpetrator belonged to the outgroup than to the ingroup. These results indicate a link between the crime and the

religious and cultural background of the outgroup perpetrator which might be interpreted as an attempt to derogate the outgroup's culture as possibly responsible for the deviant act (Khosrowtaj, Teige-Mocigemba, et al., 2023; Khosrowtaj, Yzerbyt, et al., 2023).

General Discussion

In the present work, we investigated three factors, which are discussed regarding their role in the evaluation of perpetrators from an intergroup perspective: the impact of the perpetrator's ethnicity (German ingroup, and Syrian or Afghan refugee as outgroup in Exp. 1-2), guilt certainty (Exp. 1-2) and infrahumanization (Exp. 2) in the context of sexual aggression (Experiment 1) and theft (Experiment 1-2).

Main findings regarding perpetrator and victim blaming

In two experiments, we observed the BSE (in Experiment 1 only for retributive affect used by van Prooijen, 2006, whereas we replicated the BSE for both measures in Exp. 2). In line with a BSE, participants were more likely to put the main blame on the ingroup perpetrator (relative to the victim). Participants blamed the victim more when the perpetrator stemmed from the outgroup than from the ingroup (Exp. 1-2). Thus, whereas participants indicated harsher retributive judgement towards the ingroup perpetrator, the victim was judged more harshly when the perpetrator stemmed from the outgroup. We argue that this might be considered as a shift of blame from the outgroup perpetrator to the ingroup victim while the outgroup as a whole is dehumanized and thus exonerated (Bastian et al., 2011) or culturally blamed (Khosrowtaj, Teige-Mocigemba, et al., 2023; Khosrowtaj, Yzerbyt, et al., 2023). This theorizing is indirectly supported by our finding that participants searched for stereotypic information regarding the perpetrator's background, culture, and religious affiliation when confronted with an outgroup perpetrator rather than an ingroup perpetrator (Exp. 2). Obviously, participants consider this information as differentially diagnostically relevant depending on the perpetrator's origin, possibly resulting in decreased responsibility and blame attributions. This idea is supported by the observation as part of Experiment 2, that with increasing attribution of secondary positive emotions participants were less (more) severe in their judgments towards the outgroup perpetrator (ingroup victim). Instead of leading to severe punishment of the outgroup perpetrator, attributing higher secondary positive emotions towards the outgroup exonerated the outgroup perpetrator. This pattern even though correlational is interesting as it might hint to a shift of blame away from the outgroup perpetrator to the ingroup victim (Khosrowtaj, Yzerbyt, et al., 2023).

The role of guilt certainty and infrahumanization: on some unexpected patterns

As expected, participants indicated harsher judgments towards the perpetrator in the guilt certain compared to the guilt uncertain condition (Exp. 1-2). However, based on previous literature we expected guilt certainty to moderate the occurrence of the BSE (Otten & Gordijn, 2014; van Prooijen, 2006). In Experiment 1, we observed no interaction between these factors, whereas as part of Experiment 2, we observed an interaction between guilt certainty and perpetrator's ethnicity for blame and retributive judgments. However, simple slopes revealed an unexpected pattern, namely a BSE for both the guilt certain and uncertain condition, albeit the effect was weaker within the guilt uncertain condition. Thus, we did not replicate the benefit of the doubt (Otten & Gordijn, 2014; van Prooijen, 2006) in the guilt uncertain condition which was expected to lead to harsher judgments towards the outgroup perpetrator. We do not have any explanation for this failed replication as we based our experiments on van Prooijen's (2006) work. For further understanding the pattern of findings, we added the concept of infrahumanization in Experiment 2, as we hypothesized that in some circumstances, the infrahumanization of the outgroup might override the effect of guilt certainty. We tested an experimental manipulation of infrahumanization including intergroup helping (Davies et al., 2018). We predicted that a successful decrease in infrahumanization of the outgroup might lead to equal or harsher treatment of the outgroup perpetrator in the following deviant depiction. As expected, participants in the outgroup helping the ingroup condition ascribed more positive secondary emotions to the outgroup whereas unexpectedly the ascription of secondary negative emotions was decreased in this condition. Given that our manipulation failed (specifically regarding the secondary negative emotions), we did not observe any effect of the infrahumanization manipulation on the guilt attributions. We used the ascribed emotions for investigating correlative patterns. Our analyses indicated that with increasing attribution of secondary positive emotions to Afghans, participants blamed the outgroup perpetrator (victim) less (more when the perpetrator stemmed from the outgroup).

Our results raise the question whether the valence of the emotions may be more relevant for the infrahumanization concept than has been claimed so far. The present findings are insofar interesting as past studies have often included only positive or only negative secondary emotions, making it impossible to examine differential effects (Cuddy et al., 2007; Davies et al., 2018; Demoulin et al., 2009).

However, the role of valence and human uniqueness in the infrahumanization concept has been investigated (Eyssel & Ribas, 2012; Terskova & Agadullina, 2022). Valence and

uniqueness are factors that are differentiated in lay perceptions of emotions (Terskova & Agadullina, 2022). Interestingly, positive secondary emotions were rated as more unique to humans than secondary negative emotions when asked to assess the degree to which emotions were characteristic to humans (Viki & Abrams, 2003).

Future directions

The present research hints to an important shift of blame considering the BSE and the victim blaming patterns which exonerate a dehumanized outgroup. On the other hand, our attempts to humanize the outgroup did not work out fully and require further clarification of the underlying mechanisms (e.g., the role of valence) of infrahumanization. Specifically, further research is needed to investigate whether it is possible to humanize perpetrators in intergroup deviant settings. Legal and societal consequences of infrahumanization should be investigated further for preventing discrimination of dehumanized outgroups.

One might argue that it is not clear whether participants perceived the male perpetrator and female victim as the comparison framework or the ethnicities of both parties. However, with the inclusion of theft as crime type we provided a scenario which was less confounded with gender identification and stereotypes about the involved parties (Süssenbach & Bohner, 2011; van der Bruggen & Grubb, 2014).

Conclusions

In the present work, we highlight that ingroup compared to outgroup deviants are more negatively judged. In line with this shift of blame, victim blaming was higher when the perpetrator stemmed from the outgroup. In our experiments, we did not find prove for the benefit of the doubt. That is, we did not find any interaction between guilt certainty and perpetrator ethnicity on the evaluation of the perpetrator (Exp. 1), whereas in our second experiment, both in the guilt certain and guilt uncertain condition the ingroup perpetrator was blamed more harshly compared to the outgroup perpetrator. An experimental manipulation aimed at reducing infrahumanization did not work out as predicted by previous literature. In line with the stereotypic perception of outgroups, participants indicated higher interest towards stereotypic information categories in case of the outgroup perpetrator. It remains open what drives this bias; however, the present work highlights the existence of associating deviant outgroup acts with specific causes such as religious and cultural background of the perpetrator.

References

Bastian, B., Denson, T. F., & Haslam, N. (2013). The Roles of Dehumanization and Moral Outrage in Retributive Justice. *PLOS ONE*, *8*(4), Article 4. https://doi.org/10.1371/journal.pone.0061842

Bastian, B., Laham, S. M., Wilson, S., Haslam, N., & Koval, P. (2011). Blaming, praising, and protecting our humanity: The implications of everyday dehumanization for judgments of moral status. *British Journal of Social Psychology*, *50*(3), 469–483. https://doi.org/10.1348/014466610X521383

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), 1785–1797. https://doi.org/10.1177/0886260510372945

Brems, C., & Wagner, P. (1994). Blame of Victim and Perpetrator in Rape Versus Theft. *Journal of Social Psychology*, *134*(3), 363–374. https://doi.org/10.1080/00224545.1994.9711741

Bruneau, E., Kteily, N., & Laustsen, L. (2018). The unique effects of blatant dehumanization on attitudes and behavior towards Muslim refugees during the European 'refugee crisis' across four countries. *European Journal of Social Psychology*, *48*(5), 645–662. https://doi.org/10.1002/ejsp.2357

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infrahumanization in response to collective responsibility for intergroup killing. *Journal of Personality and Social Psychology*, *90*(5), 804–818. https://doi.org/10.1037/0022-3514.90.5.804

Cuddy, A. J. C., Rock, M. S., & Norton, M. I. (2007). Aid in the Aftermath of Hurricane Katrina: Inferences of Secondary Emotions and Intergroup Helping. *Group Processes & Intergroup Relations*, *10*(1), 107–118. https://doi.org/10.1177/1368430207071344

Davies, T., Yogeeswaran, K., Verkuyten, M., & Loughnan, S. (2018). From humanitarian aid to humanization: When outgroup, but not ingroup, helping increases humanization. *PLOS ONE*, *13*(11), e0207343. https://doi.org/10.1371/journal.pone.0207343

Demoulin, S., Cortes, B., Viki, T., Rodríguez, A., Rodriguez, R., Paladino, M., & Leyens, J.-P. (2009). The Role of In-group Identification in Infra-humanization. *International Journal of Psychology : Journal International de Psychologie*, *44*, 4–11. https://doi.org/10.1080/00207590802057654

Eyssel, F., & Ribas, X. (2012). How to be good (or bad): On the fakeability of dehumanization and prejudice against outgroups. *Group Processes & Intergroup Relations*, *15*(6), Article 6. https://doi.org/10.1177/1368430212447204

Felson, R. B., & Palmore, C. (2018). Biases in blaming victims of rape and other crime. *Psychology of Violence*, *8*(3), 390–399. https://doi.org/10.1037/vio0000168

Förster, B. (2020, March 31). Syrische Familie näht 360 Masken in drei Tagen. *Potsdamer Tagesspiegel*. https://www.pnn.de/potsdam-mittelmark/coronakrise-syrische-familie-naeht-360-masken-in-drei-tagen/25698200.html

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*, 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

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), 985–1022. https://doi.org/10.1080/07418825.2018.1524507

Khosrowtaj, Z., Biermann, V., & Teige-Mocigemba, S. (2023). "He must be a foreigner" On biased information search towards in- and outgroup wrongdoers. *Manuscript Submitted*.

Khosrowtaj, Z., Teige-Mocigemba, S., & Yzerbyt, V. (2023). "(N)One of us but all of them!" Ingroup favoritism on individual and group level in the context of deviant behavior. *Manuscript Submitted*.

Khosrowtaj, Z., Yzerbyt, V., Süssenbach, P., & Teige-Mocigemba, S. (2023). On the effect of perpetrator's description on inter- and intragroup guilt attributions. *Manuscript Submitted*.

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), 901–931. https://doi.org/10.1037/pspp0000048

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Leiner, D. J. (2016). *SoSci Survey (Version 2.6.00)* [Computer software]. https://www.soscisurvey.de

Leyens, J.-P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The Emotional Side of Prejudice: The Attribution of Secondary Emotions to Ingroups and Outgroups. *Personality and Social Psychology Review*, *4*(2), 186–197. https://doi.org/10.1207/S15327957PSPR0402_06

Leyens, J.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.-P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, *31*(4), 395–411. https://doi.org/10.1002/ejsp.50

Marques, J. M., & Paez, D. (1994). The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. *European Review of Social Psychology*, *5*(1), 37–68. https://doi.org/10.1080/14792779543000011

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), 287–292. https://doi.org/10.1002/ejsp.2420180308

Maurer, M., Jost, P., Kruschinski, S., & Haßler, J. (2021). *FÜNF JAHRE MEDIENBERICHTERSTATTUNG ÜBER FLUCHT UND MIGRATION* (p. 31). Johannes Gutenberg-Universität Mainz Institut für Publizistik.

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), 165–177. https://doi.org/10.1111/spc3.12098

Reich, C. M., Pegel, G. A., & Johnson, A. B. (2021). Are Survivors of Sexual Assault Blamed More Than Victims of Other Crimes? *Journal of Interpersonal Violence*, 08862605211037423. https://doi.org/10.1177/08862605211037423

Rohmann, A., Niedenthal, P. M., Brauer, M., Castano, E., & Leyens, J.-P. (2009). The attribution of primary and secondary emotions to the in-group and to the out-group: The case of equal status countries. *The Journal of Social Psychology*, *149*(6), 709–730. https://doi.org/10.1080/00224540903348253

„Sexistisch, rassistisch, widerlich": Koblenzerin ärgert sich über AfD-Wahlplakate. (2021, February 5). https://www.rhein-zeitung.de/region/aus-den-lokalredaktionen/koblenz-und-region_artikel,-sexistisch-rassistisch-widerlich-koblenzerin-aergert-sich-ueber-afdwahlplakate-_arid,2219580.html

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

Siem, B., Mazziotta, A., Barbarino, M.-L., & Rohmann, A. (2017). „Zusammengepfercht wie die Tiere" – Der Einfluss medialer Berichterstattung auf die Entmenschlichung von Geflüchteten. *The Inquisitive Mind*, *2017*(2). https://de.in-mind.org/article/zusammengepfercht-wie-die-tiere-der-einfluss-medialer-berichterstattung-auf-die

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Muslim Immigration, Critical Events, and the Seeds of Majority Members' Support for Radical Responses: An Interactionist Perspective. *Personality and Social Psychology Bulletin*, *45*(1), 133–145. https://doi.org/10.1177/0146167218780989

Süssenbach, P. (2016). Vergewaltigungsmythen und Entscheidungen in Vergewaltigungsfällen – Eine Übersicht mit Metaanalyse. *Recht und Psychatrie*, *34*, 35–42.

Süssenbach, P., & Bohner, G. (2011). Acceptance of sexual aggression myths in a representative sample of German residents. *Aggressive Behavior*, *37*(4). https://doi.org/10.1002/ab.20390

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), 2323–2344. https://doi.org/10.1177/0886260515591975

Terskova, M. A., & Agadullina, E. (2022). The Role of Valence and Uniqueness of Emotions in the Context of Infrahumanization Theory. *Psychology in Russia: State of the Art*, *15*(1), 179–192. https://doi.org/10.11621/pir.2022.0111

van der Bruggen, M., & Grubb, A. (2014). A review of the literature relating to rape victim blaming: An analysis of the impact of observer and victim characteristics on attribution of blame in rape cases. *Aggression and Violent Behavior*, *19*(5), 523–531. https://doi.org/10.1016/j.avb.2014.07.008

van Prooijen, J.-W. (2006). Retributive reactions to suspected offenders: The importance of social categorizations and guilt probability. *Personality & Social Psychology Bulletin*, *32*(6), 715–726. https://doi.org/10.1177/0146167205284964

van Prooijen, J.-W., & Lam, J. (2007). Retributive justice and social categorizations: The perceived fairness of punishment depends on intergroup status. *European Journal of Social Psychology*, *37*(6), 1244–1255. https://doi.org/10.1002/ejsp.421

Viki, G. T., & Abrams, D. (2003). Infra-humanization: Ambivalent sexism and the attribution of primary and secondary emotions to women. *Journal of Experimental Social Psychology*, *39*(5), 492–499. https://doi.org/10.1016/S0022-1031(03)00031-3

Viki, G. T., Fullerton, I., Raggett, H., Tait, F., & Wiltshire, S. (2012). The Role of Dehumanization in Attitudes Toward the Social Exclusion and Rehabilitation of Sex Offenders. *Journal of Applied Social Psychology*, *42*(10), 2349–2367. https://doi.org/10.1111/j.1559-1816.2012.00944.x

Vollmar, A. (2021, February 15). *DIE LINKE. Koblenz fordert das Entfernen von diffamierendem AfD-Plakat*. DIE LINKE. Landesverband Rheinland-Pfalz. https://www.dielinke-rhlp.de/nc/die-linke/presse/pressemitteilungen-des-landesverbandes/detail/news/die-linke-koblenz-fordert-das-entfernen-von-diffamierendem-afd-plakat/

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual violence: 'Intersectional stereotyping' in mass media representations of male Muslim migrants in Germany. *Culture and Religion*, *20*(3), 248–271. 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), 21–47. https://doi.org/10.1177/02673231211012173

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), 257–295. https://doi.org/10.1080/14792772043000059

Zick, A. (2017). Das Vorurteil über Muslime. In P. Antes & R. Ceylan (Eds.), *Muslime in Deutschland: Historische Bestandsaufnahme, aktuelle Entwicklungen und zukünftige Forschungsfragen* (pp. 39–57). Springer Fachmedien. https://doi.org/10.1007/978-3-658-15115-7_3

Summary of Manuscript 3

Khosrowtaj, Z., Yzerbyt, V., Süssenbach, P., & Teige-Mocigemba, S. (2023). On the effect of perpetrator's description on inter- and intragroup guilt attributions. *Manuscript Submitted*.

As part of this manuscript, we examined the effect of perpetrators' ethnicity and further descriptions of the perpetrator (using the SCM, elaborated further above) on perpetrator and victim evaluation focusing on sexual offence as deviant act. So far, we have seen that participants are indeed biased regarding their interest towards stereotypical information categories but in some instances, when it comes to the evaluation of the perpetrator, the ingroup perpetrator elicited harsher hostile emotions (manuscript 2, Exp. 1) and blame judgments (manuscript 2, Exp. 2) than his outgroup counterpart. As there are other stereotypes towards groups of people such as regarding their warmth and competence, as part of this manuscript, we examined the joint effect of perpetrator's ethnicity and (no) further description of the perpetrator on blame attributions. We aimed to test whether German participants would express a black sheep effect on the one hand and exonerate a warm and competent ingroup perpetrator than his cold and incompetent counterpart. In Experiment 1 (N = 383) we conducted a 2 (perpetrator ethnicity: German vs. Afghan) × 3 (description valence: positive vs. negative vs. no further description) between-participants factorial design. We asked participants to work through an alleged newspaper article depicting either an ingroup or an outgroup perpetrator who was (not) described further as positive or negative. In Experiment 2 (N = 1048) we conducted a 2 (perpetrator ethnicity: ingroup vs. outgroup) × 2 (victim ethnicity: ingroup vs. outgroup) × 3 (description valence: no information vs. positive vs. negative) design with all factors varying between participants. As such, as part of Experiment 2, we further manipulated the victim's ethnicity for examining the full model of intra- and intergroup dyads. Results indicated in both experiments a black sheep effect, that is the ingroup perpetrator was blamed more harshly compared to his outgroup counterpart. Both experiments revealed that the positive description of the perpetrator exonerated him, that is participants blamed the negatively described perpetrator more harshly compared to his positive counterpart. Furthermore, this exoneration pattern was evident for the classification of the incident as rape. Participants perceived the incident less as rape when the perpetrator was described in positive than in negative terms. In addition to the black sheep effect on perpetrator level, we observed participants blaming the victim more harshly when the perpetrator stemmed from the

outgroup (than from the ingroup). This shift of blame from the stereotyped outgroup to the ingroup perpetrator as well as to the ingroup victim was also observed as part of manuscript 2. These results paved the path for implementing and examining the blame attributed to the culture of the involved perpetrator and victim. With this we were able to detect possible shift of blames towards the outgroup culture as a whole in case of the outgroup perpetrator while on perpetrator and victim level, those individuals from the ingroup would be blamed more for the deviant act. Furthermore, by manipulating also the victim's origin, we aimed to examine whether the outgroup victim would be blamed more harshly in an intergroup dyad, as it was the case for the ingroup victim in such a dyad. Interestingly, as part of Experiment 2, we observed again higher victim blaming when the perpetrator stemmed from the outgroup. In addition, the victim was blamed more harshly when stemming from the ingroup than the outgroup. As such, our results were not only due to the intergroup constellation of the dyads (George & Martínez, 2002): only the ingroup victim in the intergroup dyad constellation was judged more harshly. This was not the case for the outgroup victim harmed by the ingroup perpetrator. As expected, the results concerning the culture blaming indicated a higher derogation of the outgroup culture than the ingroup culture. That is, while on perpetrator level one outgroup perpetrator was exonerated (the ingroup was judged more harshly, black sheep) on cultural level, the whole cultural norms and values were at stake and responsible for the deviant act more than the ingroup culture was.

Manuscript 3

On the effect of perpetrator's description on inter- & intragroup guilt attributions

Zahra Khosrowtaj[1], Vincent Yzerbyt[3], Philipp Süssenbach[2], & Sarah Teige-Mocigemba[1]

[1]University of Marburg, Gutenbergstr. 18, 35037 Marburg, Germany

[2] Fachhochschule des Mittelstands Bielefeld

[3] University catholique de Louvain, Belgium

This research was supported by a grant from the German Research Foundation (Research Training Group 2271: "Breaking Expectations", project number: 290878970, project 11).

We wish to thank our student assistants Helena Huber, Lea Teutenberg, Sophie Gosch, Hannah Mackenrodt for their marvelous assistance in data collection. We further thank Alica Özdogan and Linda Hagen for their contribution to the preparation of the Experiments as well as their support in data collection.

Correspondence concerning this article should be addressed to Zahra Khosrowtaj. E-mail: zahra.khosrowtaj@uni-marburg.de.

Disclosure statement No potential conflict of interest was reported by the authors.

ORCID IDs

Zahra Khosrowtaj: 0000-0001-5327-9656

Vincent Yzerbyt: 0000-0003-1185-4733

Philipp Süssenbach: 0000-0001-5643-405X

Sarah Teige-Mocigemba: 0000-0002-7516-1293

Abstract

Media analyses reveal an attributional bias of outgroup crimes to Muslim religion and culture. At the same time, people tend to punish deviant ingroup members more – a phenomenon termed the black sheep effect (BSE). In the present work, we investigated the effects of perpetrators' (Exp. 1-2) and victims' (Exp. 2) ethnicity on perpetrator and victim evaluation. Moreover, we studied the impact of descriptions of the perpetrator (using the Stereotype Content Model, Exp. 1-2) on these evaluations. Experiments 1 (N = 383) and 2 (N = 1048) revealed 1) a BSE, 2) an exoneration of the warm and competent (vs. cold and incompetent) perpetrator, and 3) harsher victim blaming when the perpetrator stemmed from the outgroup. In Experiment 2, we observed that the BSE-associated exoneration of the outgroup perpetrator was associated with increased outgroup culture blaming. The incident was perceived less as rape when the perpetrator was described in positive terms (Exp. 1-2), whereas it was perceived more as rape when the perpetrator stemmed from the ingroup (Exp. 2). Future directions are discussed.

Keywords: black sheep effect, outgroup derogation, culture blaming, guilt attributions, Muslim discrimination, stereotype content model

"It is the illusion of sovereignty that makes cages of categories, i.e. the presumptuous notion that our own narrow, limited world view is complete, entire and universal, and the arrogant belief that we can wholly comprehend another person in all their complexity, that we can even wholly understand an entire constructed category of people. More than 70 million people become *the* refugee. 1.9 billion people become *the* Muslim and half the world's population becomes *the* woman. *The* Black man. *The* woman with a disability. *The* homosexual. *The* migrant worker. *The* non-binary person." (Gümüsay, 2022 p. 136)

The former German federal minister of interior, Mr. Seehofer, once called migration the "mother of all problems". Indeed, there is no doubt that terrorist and sexually violent acts in different European cities sparked heated debates about the integration of migrants and asylum seekers from predominantly Muslim countries. This issue peaked with the refugee crisis in 2015. Unsurprisingly, media analyses reveal a representational bias of naming the origin and religion of perpetrators when they are non-German (Hestermann, 2019; Kearns et al., 2018). When participants had to imagine reading a newspaper article depicting a crime and report the degree of their interest in receiving information about different categories, they expressed higher interest towards stereotypical information categories such as ethnical background and religious affiliation in face of the outgroup than ingroup perpetrator (Khosrowtaj, Biermann, et al., 2023). More often than not, references are made to Muslim culture and religion as an explanation for deviant acts such as sexual violence (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022). The impression is then that the condemnation of such events will be particularly marked. At the same time, however, ample empirical evidence suggests that people punish ingroup deviants more harshly than similarly deviant outgroup members, a phenomenon known as the black sheep effect (Marques et al., 1988; Marques & Yzerbyt, 1988). With the present work, we use a stereotyped outgroup and a relevant crime in order to shed light on the differential evaluation of ingroup and outgroup deviants. In addition, we want to examine the evaluation of the deviants' culture as a possible explanation for the deviant act. No research has yet examined attributions of blame towards individual Afghan/Muslim[10] perpetrators while accounting for the concurrent judgments of responsibility and blame attributions to the perpetrator's culture. The present work aims to fill this gap.

[10] A racialisation of the religion Islam is observed even though a nationality should not be translated into a religious affiliation; there is a synonymous use of categories such as Muslim, Turk, Arab, Migrant (Shooman, 2012) even though "not all Arabs (here: Afghans) are Muslims nor all Muslims are Arabs" (Kteily et al., 2015, p. 42).

The black sheep effect: Protecting the ingroup

The black sheep effect can be seen as a sophisticated form of ingroup favoritism (Marques et al., 1988, 1992; Marques & Yzerbyt, 1988). When confronted with deviant acts, people react with harsher punishment toward an ingroup deviant compared to an outgroup deviant, presumably to protect the ingroup as a whole. In other words, the deviant member is classified as an exception to the rule and the norms of the ingroup remain intact (Carnaghi & Yzerbyt, 2007; Marques et al., 1988; Marques & Yzerbyt, 1988; Otten & Gordijn, 2014; Yzerbyt et al., 2000). Beyond ingroup and outgroup categorization, there may be further categorization processes, however, that affect intergroup perceptions (e.g., warmth and competence of different groups).

Stereotype Content Model and the evaluation of deviant acts

Research indicates that people are seldom characterized on a single univalent dimension (Cuddy et al., 2009). The stereotype content model, SCM (for a recent review, see Abele et al., 2021; Fiske, 2018; Fiske et al., 2002) is an influential model that describes the way people perceive members of different groups comprising society. The SCM proposes that social evaluations rest on the two dimensions of warmth (friendliness, trustworthiness) and competence (capability, assertiveness), respectively associated with the groups' intent (warmth) and capability (competence). In turn, the evaluations of the groups on these two dimensions predict emotional prejudices (Fiske, 2018; Fiske et al., 2002). Ingroup members are typically perceived as warm and competent (Fiske, 2018; Fiske et al., 2002) whereas Arabs/Muslims score low on both SCM dimensions (Cuddy et al., 2009 but see; Kotzur et al., 2019). The SCM provides the theoretical background for a better understanding of how group members are perceived and what kind of behavior they are likely to receive in consequence. In the present research, our aim was not to assess warmth and competence perceptions as a dependent variable (as is often done in research based on the SCM). Rather, we were interested in the way these perceptions influence guilt attributions. In other words, we wanted to see some of the consequences of describing the perpetrator either as warm *and* competent or as cold *and* incompetent. Previous work showed relations of the status of harm doers on his evaluation (Gleason & Harris, 1976; Helmke, 2014). Defendants of an armed robbery were perceived as less blameworthy when they belonged to the middle than low socioeconomic status, which was manipulated by providing information on the defendant's educational background, occupation, and income (Gleason & Harris, 1976). Another study shed light on exonerating factors such as political

similarity to a defendant (Helmke, 2014). These findings illustrate the impact of socioeconomic status and other characteristics of a perpetrator on judges.

The present work

In the present work, we used between-participants designs to investigate whether in line with the black sheep effect, people punish an individual ingroup (German) harm-doer more harshly compared to an outgroup (Afghan) counterpart. Further, to clarify the impact of the intra- vs. intergroup nature of the situation, we investigated in how far the ethnicity of the victim (Exp. 2) influences blame attributions. Importantly, we also checked whether the outgroup's culture was more strongly derogated for the deviant act compared to the ingroup's culture (Exp. 2). Last but not least, both experiments examined the influence of positive descriptions of the harm-doer (using the stereotype content model) on participants' evaluations.

Experiment 1

Experiment 1 provides a first examination of whether German participants would produce harsher judgments towards an ingroup than an outgroup perpetrator, presumably as a means to restore the ingroup's image by symbolically excluding the deviant member (Fousiani et al., 2019; Yzerbyt et al., 2000). In addition, we tested whether different descriptions of the perpetrator would affect participants' evaluations. We built upon the stereotype content model (for a recent review, see Abele et al., 2021; Fiske, 2015; Fiske et al., 2002; Yzerbyt, 2016) and tested whether participants would more readily exonerate warm and competent perpetrators as compared to a cold and incompetent perpetrator or a not further described perpetrator. Thus, for Experiment 1, we relied on a 2 (perpetrator ethnicity: German vs. Afghan) × 3 (description valence: positive vs. negative vs. no further description) between-participants factorial design.

Methods

The data and analysis syntax that support the findings of this study are openly available in OSF at https://osf.io/rn8sm/?view_only=23b868a50c2245b8b75bc512be84d49f. For both experiments informed consent was obtained.

Sample size estimation

We estimated sample size using G*Power Version 3.1.9.2. (Faul et al., 2009). For a fixed regression model, deviation from zero with five predictors (see results section for

details regarding the multiple regression models examined), we assumed a small effect size of $f^2 = .05$, $\alpha = .05$ and power = .95. The recommended sample size was $N = 402$.

Participants

We recruited a total of $N = 425$ participants via online and public postings (e.g., mailing lists, social media). Participation lasted about fifteen minutes. We motivated participants to participate by means of a lottery including 10 gift cards (for a total value of 200€). Based on a priori specified exclusion criteria, we excluded a total of 42 participants either because they had participated in previous related experiments ($n = 4$), had a first language other than German, ($n = 24$) or did not report the mother tongue ($n = 3$) or had a at least one parent with a mother tongue of a predominantly Muslim country ($n = 11$). The final sample comprised $N = 383$ ($n = 188$ female, $n = 133$ male, $n = 5$ diverse, $n = 57$ did not indicate their gender) with an age of $M = 29.63$ years ($SD = 11.49$; range: 18 – 73 years; $n = 5$ participants did not indicate their age).

Procedure and Measures

Overall procedure

We randomly assigned participants to one of two versions of a newspaper article describing a deviant act committed by either an ingroup or outgroup perpetrator. For each version, we randomly provided participants with positive, negative, or no further description of the perpetrator (see newspaper article in supplementary materials for details). In the latter condition, only the name of the perpetrator revealed his origin (ingroup vs. outgroup). Next, participants evaluated the perpetrator and the victim. They then worked through a set of filler measures not relevant for the scope of the present experiment. Finally, participants completed demographics and had the possibility to take part in the lottery.

Newspaper article. Our first experimental manipulation, *perpetrator ethnicity*, concerned the introduction of the male perpetrator as Andreas S. or Ahmed S. (see Clark et al., 2013, for details about names as cues for races and cultures). Participants read a case of Lena, a 28-year old female who accused the perpetrator of rape.

As to our second manipulation, *description valence*, we assigned participants to a no information condition, a positive description condition, or a negative description condition. We built the descriptions by manipulating both stereotype content dimensions of warmth and competence (i.e., positive = warm and competent; negative = cold and incompetent) (Cuddy et al., 2009; Fiske et al., 2002). All participants read about the Afghan or German (Rhineland) origin of the perpetrator with an assumed religious background in case of the

outgroup perpetrator (Kauff, 2022). Those of the two information conditions also learned that police investigation had revealed additional details about the suspect (see Table 1) and that they would have a chance to read this information.

Table 1. Experimental between-subject design and descriptions of the conditions.

Perpetrator	No further description	Indicating background, positive description (warm and competent)	Indicating background, negative description (cold and incompetent)
Andreas S.	--	Andreas S., 35, was born in the Rhineland and lives in Frankfurt. He works as a dentist in his own practice. He is described as friendly and helpful by his neighbors.	Andreas S., 35, was born in the Rhineland and lives in Frankfurt. He is currently unemployed and has no completed vocational education. He is described as impolite and selfish by his neighbors.
Ahmed S.	--	The 35-year old Ahmed S. is of Afghan origin and lives in Frankfurt. He works as dentist in his own practice. He is described as friendly and helpful by his neighbors.	The 35-year-old Ahmed S. is a refugee from Afghanistan who lives in Frankfurt. He is currently unemployed and has no completed vocational education. He is described as impolite and selfish by his neighbors.

Judgments. We assessed the verdict given to the perpetrator with three items (Cronbach's α = 81): "How likely is Andreas/Ahmed guilty of a crime?" (Süssenbach, 2016; Süssenbach et al., 2012, 2017) and "How strongly should Andreas/Ahmed 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).

We measured perpetrator (Cronbach's α = .85) and victim blaming (Cronbach's α = .84) with the following six items: "How much is Andreas/Ahmed/Lena to be held responsible for what has occurred?", "How much influence did Andreas/Ahmed/Lena have on the 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 Andreas/Ahmed/Lena could have avoided the incident?" and "How much do you think Andreas/Ahmed/Lena 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 Andreas/Ahmed/Lena?" and "How hostile do you feel toward Andreas/Ahmed/Lena?" on a 7-point scale varying from 1 = 'no anger/hostility at all' to 7 = 'very much anger/hostility' (van Prooijen, 2006).

Finally, participants indicated the extent to which they classified the incident as rape (Bridges, 1991) on a 7-point scale ranging from 1 = 'definitely not rape' to 7 = 'definitely rape'.

Results

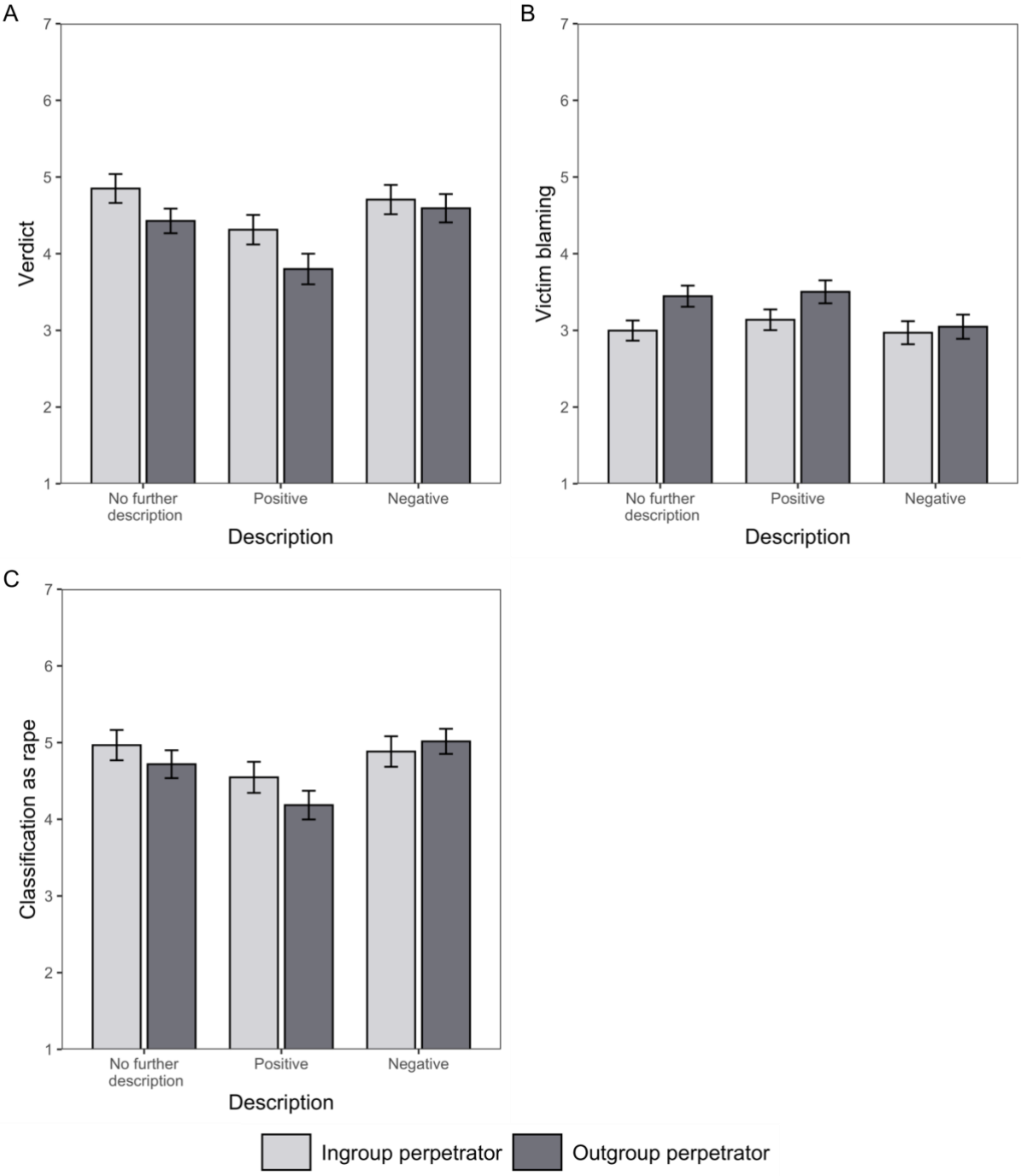
We ran a regression analysis with verdict as criterion and perpetrator ethnicity, description valence, and their interaction as predictors. Specifically, we contrast-coded perpetrator ethnicity (ingroup = - ½, outgroup = ½). As for description valence, we relied on two contrasts, namely C1, comparing the no information condition (coded 2/3) with the two valenced description conditions (each coded -1/3) and C2, comparing the positive description (½) with the negative description (-½) and 0 for the no information condition. We also included the interaction terms between perpetrator and C1 as well as between perpetrator and C2.

This analysis (see Figure 1, panel A) yielded a significant effect of perpetrator ethnicity, b = -0.35, $t(377)$ = -2.30, p = .022, η_p^2 = .01. Participants expressed a harsher verdict towards the ingroup perpetrator (M = 4.62, SD = 1.50) compared to the outgroup perpetrator (M = 4.27, SD = 1.51). There was also a significant effect of C2, b = -0.59, $t(377)$ = -3.17, p = .002, η_p^2 = .03, such that participants in the positive description condition (M = 4.05, SD = 1.59) reported a more lenient verdict towards the perpetrators compared to participants in the negative description condition (M = 4.65, SD = 1.47). There was no significant effect of C1, b = 0.29, $t(377)$ = 1.78, p = .075, nor any interaction effect: perpetrator × C1, b = -0.11, $t(377)$ = - 0.34, p = .732, perpetrator × C2, b = -0.40, $t(377)$ = -1.07, p = .287.

Next, we conducted a multiple regression, using victim blaming as criterion and the same predictors as in the previous model. This analysis revealed significant effects of perpetrator ethnicity, b = 0.30, t(377) = 2.52, p = .012, η_p^2 = .02, and C2, b = 0.31, t(377) = 2.15, p = .032, η_p^2 = .01. That is, participants blamed the victim more when the perpetrator belonged to the outgroup (M = 3.34, SD = 1.22) than when the perpetrator belonged to the ingroup (M = 3.04, SD = 1.09). In addition, mirroring the above C2 effect, participants blamed the victim less when the perpetrator was described negatively (M = 3.00, SD = 1.21) rather than positively (M = 3.32, SD = 1.16), see Figure 1, panel B. There was no significant effect of C1, b = 0.06, t(377) = 0.46, p = .646, nor any interaction effect: perpetrator × C1, b = 0.23, t(377) = 0.92, p = .361, perpetrator × C2, b = 0.29, t(377) = 0.99, p = .324.

Finally, we conducted a regression analysis with classification as rape as criterion and the same predictors as in the previous models. This analysis revealed a significant effect of C2, b = -0.58, t(377) = -3.07, p = .002, η_p^2 = .02 (see Figure 1, panel C). Participants were less likely to classify the incident as rape when the perpetrator was described positively (M = 4.36, SD = 1.57) than negatively (M = 4.95, SD = 1.42). There was no significant effect of perpetrator, b = -0.16, t(377) = -1.03, p = .303, no effect of C1, b = 0.19, t(377) = 1.13, p = .258, nor any interaction effect: perpetrator × C1, b = -0.13, t(377) = -0.41, p = .682, perpetrator × C2, b = -0.49, t(377) = -1.30, p = .194.

Figure 1: Judgments as a function of description and perpetrator's group membership



Note. A: Verdict as a function of description and perpetrator, B: Victim blaming as a function of description and perpetrator, C: Classification of the incident as rape as a function of description and perpetrator. Higher scores indicate harsher judgments. Error bars indicate standard error of the mean.

Discussion

In Experiment 1, we clearly observed a black sheep effect (Marques et al., 1988; Marques & Paez, 1994; Marques & Yzerbyt, 1988). Participants provided a harsher verdict for an ingroup perpetrator relative to an outgroup perpetrator. Also as expected, the description of the perpetrator influenced participants' verdict. Perpetrators described as warm and competent as opposed to cold and incompetent received a more lenient verdict. Moreover, participants were less likely to classify the incident as rape when the perpetrator was described in positive than in negative terms.

Turning to victim blaming, and next to the fact that participants blamed the victim less when the perpetrator was cold and incompetent rather than warm and competent, we also observed more victim blaming when the perpetrator belonged to the outgroup rather than to the ingroup. This pattern mirrors and extends the black sheep effect observed on the verdict for the perpetrator. It is also in line with effects observed for inter-ethnic rapes compared to intra-ethnic rapes, where the victim (perpetrator) is perceived as more (less) guilty (George & Martínez, 2002).

Clearly, because we did not manipulate the ethnicity of the victim, that is, the victim was always an ingroup female (German), we cannot conclude whether the inter-ethnic setting accounts for our finding. Indeed, the more lenient judgments of the outgroup perpetrator as compared to the ingroup one may also contribute to expressing harsher judgments on the ingroup victim. In other words, there might be a shift of blame when the outgroup perpetrator is exonerated. This is all the more likely as the racialization and islamisation of sexual violence observed in German print media (Wigger, 2019; Wigger et al., 2022) might lead to the prejudicial assumption that the outgroup as a whole does behave non-morally (that is, sexually aggressive and misogynic). An intriguing aspect of this rationale is that the outgroup perpetrator is not so much condemned as the entire outgroup culture is. Put differently, and in line with a broader interpretation of the black sheep effect, the ingroup perpetrator is blamed precisely because the ingroup culture is viewed as alien to such deviant behavior. In keeping with this rationale, a complementary story is then likely to prevail for the outgroup perpetrator as the exoneration of the perpetrator vis-à-vis the blaming of the victim might be a consequence of blaming the outgroup and its culture and norms as a whole.

Experiment 2 tested this conjecture more directly and had two aims. As in Experiment 1, we again examined the impact of an additional positive or negative description. Moreover, we also manipulated the victim's origin to create both intra- and inter-ethnic

dyads of perpetrator and victim. This allowed us to test whether the pattern observed for a possible outgroup victim replicates the findings by George and Martínez (2002). Second, we included a measure of culture blaming to check whether a shift of blame toward the outgroup culture as a whole emerges in case of an outgroup perpetrator while perceivers judge the outgroup perpetrator more leniently and they blame the ingroup perpetrator and victim.

Experiment 2

In Experiment 2, we hoped to replicate a) the black sheep effect, b) the exoneration of the positively described perpetrator, c) harsher victim blaming in case of an outgroup perpetrator, and d) the positive description decreasing the likelihood of perceiving the incident as rape. With respect to the attributions of responsibility to the culture of the protagonists, we predicted that participants would assign higher responsibility to the outgroup culture than to the ingroup culture. We tested a 2 (perpetrator ethnicity: ingroup vs. outgroup) × 2 (victim ethnicity: ingroup vs. outgroup) × 3 (description valence: no information vs. positive vs. negative) design with all factors varying between participants.

Methods

Sample size estimation

We estimated sample size as in Experiment 1 using G*Power Version 3.1.9.2. (Faul et al., 2009). We relied on a fixed regression model, deviation from zero with eleven predictors (see results section for details regarding the multiple regression models examined) and a power of 0.95. We calculated f^2 (R^2 / 1-R^2) using the effects observed in experiment 1 for verdict (R^2 = .019, f^2 = .0194 indicating a N = 1103) and victim blaming (R^2 = .029, f^2 = .0298 indicating a N = 854). As we did not know the effect size for the culture blaming measure, we aimed for N = 1103.

Participants

We recruited a total of N = 1126 participants via online and public postings (e.g., mailing lists, social media). We motivated participants to participate by means of a lottery including 10 gift cards (for a total value of 200€). We excluded 78 participants in total. As in Experiment 1, participants were excluded because of a first language other than German (n = 38) or no indication regarding their first language (n = 6). We further dropped participants with at least one parent with a mother tongue of a predominantly Muslim country (n = 24) as well as those who participated in previous similar experiments (n = 5). In addition, we excluded participants with a conspicuous age (e.g., ‘0’ or ‘-29’) and those

younger than 18 years (n = 5). Thus, the following analyses are based on a total sample size of N = 1048 (n = 716 female, n = 289 male, n = 17 diverse, n = 26 did not indicate their gender) with an average age of M = 30.36 years (SD = 13.03; range: 18 – 82 years; n = 2 participants did not indicate their age).

Procedure and measures

Overall procedure

We kept Experiment 2 parallel to Experiment 1 but with the addition of another between-participants factor, namely the victim's ethnicity. We kept the other factors as described in Experiment 1.

Newspaper article. We used the same articles used in Experiment 1 with the introduction of Latifa as an Afghan outgroup victim complementing the intra- and interethnic design.

Judgments. Complementary to our measures of Experiment 1, namely verdict (3 items, Cronbach's Alpha = .80), victim blaming (6 items, Cronbach's Alpha = .86), the classification as rape item, we added a series of items for assessing the blame attributed to the culture of the perpetrator (4 items, Cronbach's Alpha = .93) and the victim (4 items, Cronbach's Alpha = .92). Two exemplary items read as follows: "(Name of the perpetrator/victim)'s behavior can in part be explained by cultural norms and values."; "How much influence did the (name of the perpetrator/victim)'s culture have on the outcome of the situation?" on 7-point scales ranging from 1 = "no influence at all" to 7 = "very strong influence".

Results

We first ran a regression analysis with verdict as criterion and with perpetrator and victim ethnicity (each contrast coded with -½ for ingroup and ½ for outgroup) as well as the two contrasts C1 and C2 for the description coded as in Experiment 1 and all interaction terms as predictors (see Figure 2, panel A). This analysis yielded a significant effect of perpetrator ethnicity, b = -0.34, t(1036) = -3.65, p < .001, η_p^2 = .013. As in Experiment 1, participants reported harsher verdicts towards the ingroup perpetrator (M = 4.57, SD = 1.48) compared to the outgroup perpetrator (M = 4.24, SD = 1.56). Also replicating Experiment 1's findings, there was a significant effect of C2, b = -0.43, t(1036) = -3.78, p < .001, η_p^2 = .014, such that participants were more lenient when the perpetrator was described in positive terms (M = 4.15, SD = 1.59) rather than in negative terms (M = 4.58, SD = 1.44). These main effects were qualified by a significant perpetrator ethnicity × victim ethnicity

× C2 three-way interaction, b = -0.96, t(1036) = -2.10, p = .036, η_p^2 = .004. No other effect was observed, all non-significant effects are depicted as part of the appendix A.

To probe this interaction, we first looked at the perpetrator ethnicity × C2 interaction as a function of the victim. This interaction term was not significant for the ingroup victim, p = .524. Interestingly, this same interaction was significant for the outgroup victim, Latifa, perpetrator ethnicity × C2, b = -0.76, t(1036) = -2.34, p < .020, η_p^2 = .005. Follow-up simple slope analyses as a function of perpetrator ethnicity showed a significant effect of C2, b = -0.67, t(1036) = -2.99, p = .003, η_p^2 = .009, in case of the outgroup perpetrator, whereas no difference between the negative and positive description condition (C2) emerged for the ingroup perpetrator, p = .717. Thus, in the case of an outgroup victim, participants proved more lenient towards the outgroup perpetrator when he was described as positive rather than negative whereas no difference emerged between the description conditions for the ingroup perpetrator.

Next, we ran a multiple regression analysis with victim blaming as criterion and included the same predictors as before. This analysis revealed significant main effects for perpetrator ethnicity, b = 0.24, t(1036) = 3.27, p = .001, η_p^2 = .010, and for victim ethnicity, b = -0.23, t(1036) = -3.08, p = .002, η_p^2 = .009. Participants blamed the victim more when the perpetrator belonged to the outgroup (M = 3.34, SD = 1.20) rather than to the ingroup (M = 3.11, SD = 1.72). Further, participants attributed harsher blame towards the ingroup victim (M = 3.34, = 1.17) than to the outgroup victim (M = 3.11, SD = 1.21) (see Figure 2 panel B).

As in Experiment 1, we also conducted a regression analysis with classification of the incident as rape as criterion. This analysis revealed a significant effect of perpetrator ethnicity, b = -0.30, t = -3.10, p = .002, η_p^2 = .009, and a significant effect of C2, b = -0.40, t = -3.34, p = .001, η_p^2 = .011. Participants rated the incident more as rape when the perpetrator belonged to the ingroup (M = 4.85, SD = 1.52) than to the outgroup (M = 4.56, SD = 1.61). Participants were less likely to rate the incident as a rape in the condition featuring a positive description of the perpetrator (M = 4.49, SD = 1.63) than in the condition featuring a negative description of the perpetrator (M = 4.87, SD = 1.47) (see Figure 2, panel C).

To examine our hypothesis regarding culture blaming, we ran a multiple regression model with blame attributed to the culture of the perpetrator as criterion and with the same predictors as before (see Figure 2, panel D). This analysis revealed a significant effect of

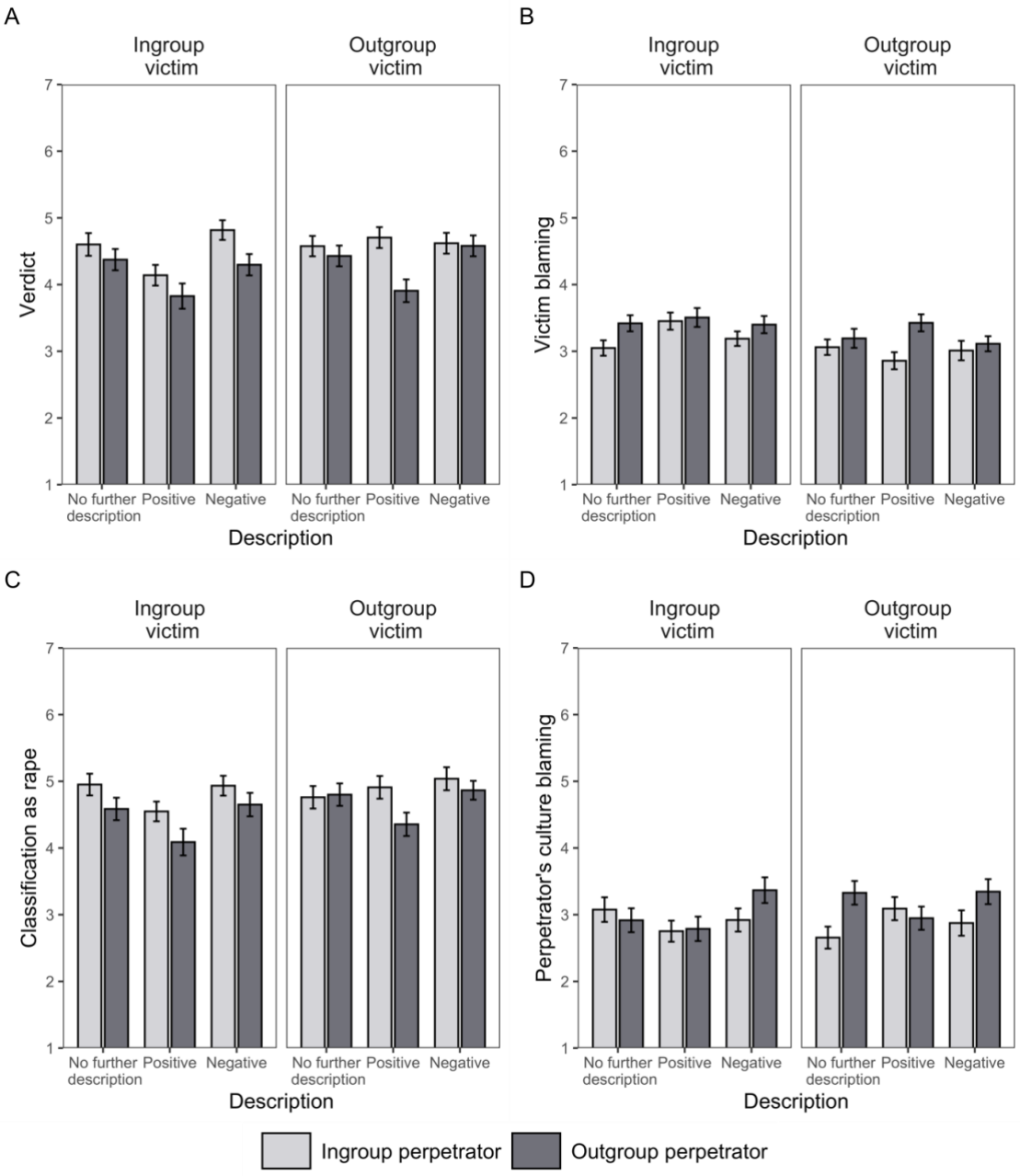
perpetrator ethnicity, b = 0.22, t(1036) = 2.14, p = .033, η_p^2 = .004. In line with our hypotheses, participants assigned more blame to the outgroup culture (M = 3.11, SD = 1.71) than to the ingroup culture (M = 2.89, SD = 1.63). There was also a significant perpetrator ethnicity × C2 interaction, b = -0.51, t(1036) = -2.02, p = .044, , η_p^2 = .004, and a significant perpetrator ethnicity × victim ethnicity × C1 interaction, b = 0.91, t(1036) = 2.09, p = .037, η_p^2 = .004.

Regarding the perpetrator ethnicity × C2 interaction, participants perceived the outgroup's culture to be more responsible when the outgroup perpetrator was described as negative, C2, b = -0.49, t(1036) = -2.71, p = .007, η_p^2 = .007, whereas this C2 effect was absent for Andreas, p = .892.

To probe the three-way interaction, we looked at the perpetrator ethnicity × C1 interaction as a function of victim ethnicity. These separate analyses for the victims did not indicate a significant effect of perpetrator × C1, p = .191, for Lena nor for Latifa, perpetrator × C1, p = .101.

As regards blame attributed to the culture of the victim, we ran a multiple regression analysis with the same predictors as before. The full regression model failed to reach significance, F(11, 1036) = 1.58, p = .100. Interestingly, however, there was an effect of victim's ethnicity, b = -0.27, t(377) = -2.83, p = .005, that is, participants assigned more blame to the victim's culture, when the victim stemmed from the outgroup (M = 4.78, SD = 1.56) than from the ingroup (M = 4.63, SD = 1.58).

Figure 2: Judgments as a function of description and the group membership of perpetrator and victim



Note. A: Verdict as a function of description, perpetrator, and victim, B: Victim blaming as a function of description, perpetrator, and victim, C: Classification of the incident as rape as a function of description, perpetrator, and victim. D: Perpetrator's culture blaming as a function of description, perpetrator, and victim. Higher scores indicate harsher judgments. Error bars indicate standard error of the mean.

Discussion

As expected, Experiment 2 replicated the black sheep effect observed in Experiment 1. We also replicated the exoneration of the perpetrator when he was described in positive terms. We further observed that the harsher verdict for the ingroup perpetrator (the black sheep effect) was more pronounced when his description was negative, and the victim belonged to the ingroup.

Regarding victim blaming, we replicated the pattern of harsher victim blaming when the perpetrator stemmed from the outgroup independently of whether the victim was an ingroup member (interethnic rape) or an outgroup member (intraethnic rape). In addition, participants blamed the victim more when she belonged to the ingroup than to the outgroup regardless of the perpetrator's ethnicity. Thus, we did not observe an interaction between perpetrator and victim ethnicity (but see George & Martínez, 2002 for stronger victim blaming in interethnic contexts as compared to intraethnic contexts). This finding doves well with the pattern observed by De keersmaecker and Roets (2020) who found a relationship between the belief in a just world and harsher ingroup victim blaming (in the context of a robbery and physical violence). One possible account is the increased similarity and threat to the ingroup when the victim is an ingroup member. Presumably, this would then motivate increased blaming of the ingroup victims for their misfortune (De keersmaecker & Roets, 2020).

Moreover, participants perceived the incident more as a rape when the perpetrator belonged to the ingroup than to the outgroup. The positive description of the perpetrator, as in Experiment 1, also decreased the likelihood of perceiving the incident as a rape compared to the negative description of the perpetrator.

In Experiment 2, we additionally looked at the blame attributed to the culture of the perpetrator and the victim, respectively. Expectedly, the outgroup culture was blamed more harshly when the outgroup perpetrator was described in negative terms and the victim belonged to the ingroup. More importantly, however, and in line with our prediction, participants assigned more blame towards the culture of the outgroup perpetrator than towards the culture of the ingroup perpetrator. The regression model on victim's culture blaming did not reach significance.

General discussion

In the present work, we investigated whether in line with the black sheep effect German participants would judge an ingroup perpetrator more harshly than an outgroup

perpetrator. We additionally examined the effect of the perpetrator's warmth and competence on participants' evaluations. Complementing the judgments about the individual perpetrator, we further assessed participants' blame attributions towards the culture of the in- and outgroup perpetrators and victims.

In both experiments, we observed a black sheep effect. Mirroring the black sheep effect observed for perpetrators, both studies found harsher victim blaming when the perpetrator stemmed from the outgroup. Also, regardless of the perpetrator's ethnicity, participants indicated more blame toward an ingroup victim than toward an outgroup victim (Exp. 2). Considering the judgments regarding perpetrator and victim this pattern constitutes a shift of blame (see also Khosrowtaj, Süssenbach, et al., 2023). Indeed, whereas the outgroup perpetrator was less blamed than the ingroup perpetrator, the victim received harsher judgments when the perpetrator stemmed from the outgroup, hinting to a shift of blame.

Going beyond the assignment of blame to individuals, Experiment 2 tested our hypothesis regarding culture blaming. In line with predictions, the data not only replicated the pattern of blame attributed to the individual level but also revealed that participants perceived the outgroup's cultural norms and values as being responsible for the deviant act when the perpetrator belonged to the outgroup. We elaborate on this pattern below.

Regarding the effects of a positive versus negative description of the perpetrator, the results from Experiments 1 and 2 are clearcut. First, a positive description of the perpetrator was associated with greater exoneration of the perpetrator. Second, participants perceived the incident more as rape when the perpetrator was described in negative than in positive terms.

The exoneration of the positively characterized (warm and competent) perpetrator (independent of his group membership) is in line with prior research showing that the success and power of a perpetrator affect labeling the incident as rape as well as judging the rape (Nyúl et al., 2018). The fact that the status of the harm doer affects the blameworthiness of a perpetrator (Gleason & Harris, 1976) may have severe consequences for victims of sexual violence such as normalizing the violence and perceiving the victim as complicit of the incident (cf. Thapar-Björkert & Morgan, 2010).

Regarding the classification of the incident as rape, participants perceived the incident more as rape when the perpetrator stemmed from the ingroup in line with the BSE (Experiment 2). Alarmingly, the positive (than negative) description had an exonerating effect as it decreased the likelihood of perceiving the incident as rape (Experiments 1-2).

Limitations and Future Directions

With regard to possible limitations, it has to be noted that in the present research, the framework of the stereotype content model served only as an example of two possible dimensions that may affect the perception of perpetrator and victim in the present context. We did not intend to investigate all possibly relevant dimensions but rather chose to focus on two dimensions, warmth and competence, that have been identified as highly relevant in research on person and group perception (Fiske, 2018). Also we did not test possible differential effects due to, for instance, ambivalent stereotypes (Fiske et al., 2002; Koch et al., 2016; Yzerbyt, 2016). Future research might disentangle the role of different stereotype dimensions as well as possible compensating effects (Kervyn et al., 2009; Yzerbyt et al., 2005) or even use other dimensions of person descriptions in the context of intra- and intergroup crime evaluations. Further, beyond the investigated crime, future research may extend this work to other deviant acts. In addition, it may be promising to look at different ethnicity groups for disentangling similarities and discrepancies between ethnicity groups.

Next to the lessons learned at the more basic level, the finding that certain perpetrator characteristics are associated with an exoneration of the deviant act is also of practical relevance. To the extent that a positive description decreased the likelihood of perceiving the deviant act as rape, this may help better understand why sexual assaults are rarely sentenced appropriately or often even go unreported. Indeed, victims may anticipate that others will not believe them if the perpetrator is generally seen in a positive light.

The combined findings regarding the black sheep effect and culture blaming clearly go beyond the previous research on the black sheep effect and encourage us to suggest that these two aspects go hand in hand. As participants indicate harsher judgments towards the ingroup perpetrator, they are also derogating the outgroup culture as a whole (Khosrowtaj, Teige-Mocigemba, et al., 2023). This pattern suggests that the outgroup culture is perceived as being somehow collectively responsible for the deviant act (Doosje et al., 2007; Kardos et al., 2019; Lickel et al., 2003). In other words, a harsher judgment of the ingroup perpetrator combined with blaming the outgroup culture is likely used by participants to signal that they consider the ingroup as being more alien to such deviance than the outgroup is. New as it stands, this message is entirely consistent with the initial take on the black sheep effect as a way to promote a positive image of the ingroup (Marques & Yzerbyt, 1988).

Conclusions

The present work sheds light on the influence of a perpetrator's ethnicity and description (warmth and competence) on guilt attributions. An alarming finding is the fact that positive (compared to negative) descriptions exonerated harm doers even though the crime was exactly the same (cf. Gleason & Harris, 1976; Nyúl et al., 2018). In contrast to media debates and despite prevalent negative attitudes and hostility towards Muslims (Bauer & Hannover, 2020; Stürmer et al., 2019; Zick et al., 2011), our work illustrated experimentally that German participants evaluate an ingroup perpetrator more harshly than an outgroup counterpart. While people judged the outgroup perpetrator more leniently, they also perceived the norms and values of the outgroup, that is, its culture, as more responsible for the deviant act. Being more lenient towards an outgroup perpetrator (Braun & Gollwitzer, 2012) or being more punitive towards an ingroup perpetrator (Marques et al., 1988; Yzerbyt et al., 2000) may thus go hand in hand with the derogation of the perpetrator's outgroup as a whole (Doosje et al., 2007; Khosrowtaj, Teige-Mocigemba, et al., 2023; Lickel et al., 2003; Stürmer et al., 2019).

Thus, the present work hints to a possible convergence of the condemnation of one individual perpetrator and the harsher derogation of the outgroup culture which in the latter case has consequences for a whole homogenized group.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

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

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

Braun, J., & Gollwitzer, M. (2012). Leniency for out-group offenders. *European Journal of Social Psychology*, *42*(7), 883–892. https://doi.org/10.1002/ejsp.1908

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

Clark, J. W., Cramer, R. J., Percosky, A., Rufino, K. A., Miller, R. S., & Johnson, S. M. (2013). Juror perceptions of African American- and Arabic-named victims. *Psychiatry, Psychology and Law*, *20*(5), 781–794. https://doi.org/10.1080/13218719.2012.736283

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., ... Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), Article 1. https://doi.org/10.1348/014466608X314935

De keersmaecker, J., & Roets, A. (2020). All victims are equally innocent, but some are more innocent than others: The role of group membership on victim blaming. *Current Psychology*, *39*(1), 254–262. https://doi.org/10.1007/s12144-017-9763-9

Doosje, B., Zebel, S., Scheermeijer, M., & Mathyi, P. (2007). Attributions of Responsibility for Terrorist Attacks: The Role of Group Membership and Identification. *International Journal of Conflict and Violence (IJCV)*, *1*(2), Article 2. https://doi.org/10.4119/ijcv-2749

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fiske, S. T. (2015). Intergroup Biases: A Focus on Stereotype Content. *Current Opinion in Behavioral Sciences*, *3*, 45–50. https://doi.org/10.1016/j.cobeha.2015.01.010

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), Article 6. https://doi.org/10.1037/0022-3514.82.6.878

Fousiani, K., Yzerbyt, V., Kteily, N.-S., & Demoulin, S. (2019). Justice reactions to deviant ingroup members: Ingroup identity threat motivates utilitarian punishments. *British Journal of Social Psychology*, *58*(4), 869–893. https://doi.org/10.1111/bjso.12312

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), 110–119. https://doi.org/10.1111/1471-6402.00049

Gleason, J. M., & Harris, V. A. (1976). Group Discussion and Defendant's Socio-Economic Status as Determinants of Judgments by Simulated Jurors1. *Journal of Applied Social Psychology*, *6*(2), Article 2. https://doi.org/10.1111/j.1559-1816.1976.tb01323.x

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

Helmke, S. (2014). *Beliefs About the Strauss-Kahn Case in France and Germany: Political Orientation and Sexual Aggression Myths as Local Versus Global Predictors*. *8*, 15.

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

Kardos, P., Leidner, B., Castano, E., & Lickel, B. (2019). The benefits of collective responsibility: How ingroup reputation concern motivates prosociality in intergroup contexts. *European Journal of Social Psychology*, *49*(1), 93–109.

https://doi.org/10.1002/ejsp.2506

Kauff, M. (2022). Muslim = Terrorist? Attribution of violent crimes to terrorism or mental health problems depend on perpetrators' religious background. *The Journal of Social Psychology*, *0*(0), 1–9. https://doi.org/10.1080/00224545.2022.2095968

Kearns, E., Betus, A., & Lemieux, A. (2018). *Why Do Some Terrorist Attacks Receive More Media Attention Than Others?* (SSRN Scholarly Paper ID 2928138). Social Science Research Network. https://papers.ssrn.com/abstract=2928138

Kervyn, N., Judd, C. M., & Yzerbyt, V. Y. (2009). You want to appear competent? Be mean! You want to appear sociable? Be lazy! Group differentiation and the compensation effect. *Journal of Experimental Social Psychology*, *45*(2), 363–367. https://doi.org/10.1016/j.jesp.2008.08.006

Khosrowtaj, Z., Biermann, V., & Teige-Mocigemba, S. (2023). "He must be a foreigner" On biased information search towards in- and outgroup wrongdoers. *Manuscript Submitted*.

Khosrowtaj, Z., Süssenbach, P., & Teige-Mocigemba, S. (2023). The black sheep as a shift of blame? Blaming the ingroup perpetrator and victim in an intergroup deviant setting. *Manuscript Submitted*.

Khosrowtaj, Z., Teige-Mocigemba, S., & Yzerbyt, V. (2023). "(N)One of us but all of them!" Ingroup favoritism on individual and group level in the context of deviant behavior. *Manuscript Submitted*.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Kotzur, P. F., Friehs, M.-T., Asbrock, F., & Zalk, M. H. W. van. (2019). Stereotype content of refugee subgroups in Germany. *European Journal of Social Psychology*, *0*(0), Article 0. https://doi.org/10.1002/ejsp.2585

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), 901–931. https://doi.org/10.1037/pspp0000048

Lickel, B., Schmader, T., & Hamilton, D. L. (2003). A Case of Collective Responsibility: Who Else Was to Blame for the Columbine High School Shootings? *Personality and Social Psychology Bulletin*, *29*(2), Article 2. https://doi.org/10.1177/0146167202239045

Marques, J. M., & Paez, D. (1994). The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. *European Review of Social Psychology*, *5*(1), Article 1. https://doi.org/10.1080/14792779543000011

Marques, J. M., Robalo, E. M., & Rocha, S. A. (1992). Ingroup bias and the 'black sheep' effect: Assessing the impact of social identification and perceived variability on group judgements. *European Journal of Social Psychology*, *22*(4), 331–352. https://doi.org/10.1002/ejsp.2420220403

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

Nyúl, B., Kende, A., Engyel, M., & Szabó, M. (2018). Perception of a Perpetrator as a Successful Person Predicts Decreased Moral Judgment of a Rape Case and Labeling it as Rape. *Frontiers in Psychology*, *9*, 2555. https://doi.org/10.3389/fpsyg.2018.02555

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

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). Muslim Immigration, Critical Events, and the Seeds of Majority Members' Support for Radical Responses: An Interactionist Perspective. *Personality and Social Psychology Bulletin*, *45*(1), 133–145. https://doi.org/10.1177/0146167218780989

Süssenbach, P. (2016). Vergewaltigungsmythen und Entscheidungen in Vergewaltigungsfällen – Eine Übersicht mit Metaanalyse. *Recht und Psychatrie*, *34*, 35–42. https://psychiatrie-verlag.de/product/suessenbach-p-vergewaltigungsmythen-und-entscheidungen-in-vergewaltigungsfaellen-

einzelartikel-aus-rp-1-2016/

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

Thapar-Björkert, S., & Morgan, K. J. (2010). "But Sometimes I Think . . . They Put Themselves in the Situation": Exploring Blame and Responsibility in Interpersonal Violence. *Violence Against Women*, *16*(1), 32–59. https://doi.org/10.1177/1077801209354374

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

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual violence: 'Intersectional stereotyping' in mass media representations of male Muslim migrants in Germany. *Culture and Religion*, *20*(3), 248–271. 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

Yzerbyt, V. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, *11*, 90–95. https://doi.org/10.1016/j.copsyc.2016.06.009

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

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory Stereotypes in the French-speaking World. *Group Processes & Intergroup Relations*, *8*(3), 291–308. https://doi.org/10.1177/1368430205053944

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

Appendix A

This appendix depicts all results including the non-significant ones. Note, that all other results (for instance for Exp. 1) have been reported as part of the results section.

Table 1. Predicting verdict based on perpetrator's and victim's group membership and perpetrator's characterization.

	b	SE	*t*	*p*
Constant	4.41	.05	95.54	< .001
Perpetrator (ingroup = - ½, outgroup = ½)	- 0.34	.09	- 3.65	< .001
Victim (ingroup = - ½, outgroup = ½)	0.13	.09	1.36	= .174
C1: comparing no further information (2/3) with description conditions (each -1/3)	0.13	.10	1.36	= .175
C2: comparing positive description (1/2) with negative description (-1/2) condition (no information condition = 0)	- 0.43	.11	- 3.78	< .001
Perpetrator × Victim	0.03	.19	0.14	= .891
Perpetrator × C1	0.23	.20	1.17	= .244
Perpetrator × C2	-0.28	.23	- 1.20	= .230
Victim × C1	-0.17	.20	- 0.85	= .394
Victim × C2	0.28	.23	1.22	= .223
Perpetrator × Victim × C1	0.08	.39	0.21	= .834
Perpetrator × Victim × C2	-0.96	.46	- 2.10	= .036

Notes. $R = .194$, $R^2 = .038$, $R^2_{corrected} = .028$, $F(11,1036) = 3.69$, $p < .001$

Table 2. Predicting victim blaming based on perpetrator's and victim's group membership and characterization.

	b	SE	*t*	*p*
Constant	3.22	.04	87.98	< .001
Perpetrator (ingroup = - ½, outgroup = ½)	0.24	.07	3.27	= .001
Victim (ingroup = - ½, outgroup = ½)	-0.23	.07	-3.08	= .002
C1: comparing no further information (2/3) with description conditions (each -1/3)	-0.06	.08	-0.82	= .412
C2: comparing positive description (1/2) with negative description (-1/2) condition (no information condition = 0)	0.13	.09	1.47	= .141
Perpetrator × Victim	0.06	.15	0.38	= .705
Perpetrator × C1	0.02	.16	0.12	= .908
Perpetrator × C2	0.16	.18	0.86	= .390
Victim × C1	0.18	.16	1.15	= .250
Victim × C2	-0.10	.18	-0.58	= .563
Perpetrator × Victim × C1	-0.44	.31	-1.42	= .156
Perpetrator × Victim × C2	0.62	.36	1.73	= .085

Notes. $R = .169$, $R^2 = .029$, $R^2_{\text{corrected}} = .018$, $F(11,1036) = 2.78$, $p = .001$

Table 3. Predicting classification of the incident as rape based on perpetrators group membership and characterization.

	b	SE	*t*	*p*
Constant	4.71	.05	97.48	< .001
Perpetrator (ingroup = - ½, outgroup = ½)	-0.30	.10	-3.10	= .002
Victim (ingroup = - ½, outgroup = ½)	0.16	.10	1.67	= .094
C1: comparing no further information (2/3) with description conditions (each -1/3)	0.10	.10	0.99	= .324
C2: comparing positive description (1/2) with negative description (-1/2) condition (no information condition = 0)	-0.40	.12	-3.34	= .001
Perpetrator × Victim	0.14	.19	0.73	= .467
Perpetrator × C1	0.21	.20	1.00	= .316
Perpetrator × C2	-0.28	.24	-1.17	= .242
Victim × C1	-0.23	.20	-1.10	= .270
Victim × C2	0.15	.24	0.65	= .516
Perpetrator × Victim × C1	0.40	.41	0.97	= .331
Perpetrator × Victim × C2	-0.20	.48	-0.43	= .668

Notes. $R = .165$, $R^2 = .027$, $R^2_{corrected} = .017$, $F(11,1036) = 2.63$, $p = .003$

Table 4. Predicting culture blaming towards perpetrator based on perpetrator's and victim's group membership and characterization.

	b	SE	*t*	*p*
Constant	3.01	.05	58.28	< .001
Perpetrator (ingroup = - ½, outgroup = ½)	0.22	.10	2.14	= .033
Victim (ingroup = - ½, outgroup = ½)	0.07	.10	0.68	= .496
C1: comparing no further information (2/3) with description conditions (each -1/3)	-0.02	.11	-0.15	= .879
C2: comparing positive description (1/2) with negative description (-1/2) condition (no information condition = 0)	-0.23	.13	-1.83	= .067
Perpetrator × Victim	0.23	.21	1.09	= .275
Perpetrator × C1	0.05	.22	0.25	= .805
Perpetrator × C2	-0.51	.25	-2.02	= .044
Victim × C1	-0.11	.22	-0.51	= .608
Victim × C2	0.28	.25	1.11	= .266
Perpetrator × Victim × C1	0.91	.44	2.09	= .037
Perpetrator × Victim × C2	-0.20	.51	-0.40	= .690

Notes. $R = .137$, $R^2 = .019$, $R^2_{\text{corrected}} = .008$, $F(11,1036) = 1.80$, $p < .049$

Supplementary materials

Experiment 1 & 2: Newspaper articles & descriptions

Frankfurt am Main

Mutmaßliche Vergewaltigung: Polizei verhört Ahmed / Andreas S.

21. Mai 2018, 11:32 Uhr

Am vergangenen Freitag soll laut Polizeibericht die 28-jährige Lena/Latifa K. vergewaltigt worden sein. Die Studentin hatte den Tatverdächtigen Ahmed/Andreas S. Anfang April über eine Dating-App kennengelernt. In der vergangenen Woche kam es zu einem Treffen in der Wohnung des Mannes. Dabei wurden geringfügige Mengen Alkohol konsumiert.

Als die junge Frau die letzte S-Bahn verpasste, soll ihr der Beschuldigte angeboten haben, auf der Couch zu übernachten. In dieser Nacht soll der 35-Jährige die Frankfurterin/Afghanin vergewaltigt haben. Lena/Latifa K. wandte sich am darauffolgenden Tag an die Polizei.

Als der Beschuldigte Ahmed/Andreas S. von der Polizei zum Verhör abgeholt wurde, zeigte er sich sichtlich überrascht und bestritt die Vergewaltigungsvorwürfe. Mit Lena/Latifa K. sei es zwar zu sexuellen Handlungen in seiner Wohnung gekommen, diese sollen jedoch einvernehmlich stattgefunden haben.

English translation:

Frankfurt am Main

Suspected rape: Police interrogates Ahmed/ Andreas S.

May 21, 2018, 11:32 a.m.

According to a police report, the 28-year-old Lena/Latifa K. was raped last Friday. The student had met the suspect Ahmed/ Andreas S. in early April through a dating app. Last week there was a meeting at the man's apartment. Minor amounts of alcohol were consumed.

When the young woman missed the last suburban train, the suspect offered her to spend the night on the couch. On this night the 35-year-old is said to have raped the woman from Frankfurt / Afghanistan. Lena/ Latifa K. contacted the police the following day.

When the accused Ahmed/ Andreas S. was picked up by the police for interrogation, he appeared visibly surprised and denied the rape allegations. Although sexual acts with Lena/Latifa K. had taken place in his apartment, they were allegedly consensual.

Summary of Manuscript 4

Khosrowtaj, Z., Teige-Mocigemba, S., & Yzerbyt, V. (2023). "(N)One of us but all of them!" Ingroup favoritism on individual and group level in the context of deviant behavior. *Manuscript Submitted*.

Following up on the previous manuscript, we conducted another set of Experiment to investigate ingroup protection on individual and cultural level more in depth. Recall that as part of the previous manuscript, we observed on the one hand, harsher judgments of a German perpetrator compared to his Afghan counterpart, and on the other hand harsher derogation of the outgroup's (than ingroup's) culture in the context of intergroup violence. The other side of the coin would translate to more lenient judgments of the ingroup on cultural level (protecting the ingroup as a whole) and condemning one individual ingroup member (again protecting the ingroup by this distancing strategy). As part of this contribution, we tested ingroup protection on individual and cultural level more in depth. We also examined some boundary conditions such as the effect of the mere presence of the outgroup category before participants learned about the crime (Exp. 1), the role of the comparison context (Exp. 2-3) and the impact of ingroup entitativity on the judgments on individual and on cultural level (Exp. 3).

In Experiment 1 (N = 437) we did not find a black sheep effect probably due to the ambiguity of the guilt (Otten & Gordijn, 2014; van Prooijen, 2006). The priming manipulation did not have an impact on the proceeding judgments probably due to being too broad (Ledgerwood & Chaiken, 2007). However, we replicated the more lenient judgments towards the ingroup than outgroup culture. Experiment 2 built upon these limitations, and we created newsflashes where the guilt of the perpetrator was certain. Further we used a within-subjects design for having a clear intergroup context. This would allow participants to perceive an intergroup comparison, especially when the ingroup newsflash is presented second.

In Experiments 2 (N = 283) we observed the predicted protection of the ingroup both on individual and on cultural level. That is, we observed the overall black sheep effect affecting the individual perpetrator and again the ingroup culture was judged more leniently than the outgroup culture. Interestingly, the black sheep effect was manifested when the outgroup newsflash was presented first. This pattern is in line with the priming idea of Exp. 1: when the outgroup category is salient, participants express ingroup protection in form of the black sheep effect. In addition, participants who began with the ingroup first expressed

harsher judgments towards the outgroup perpetrator who came second. For testing the robustness of these order effects, we followed up with Experiment 3. We expected to replicate specially the black sheep effect when the outgroup newsflash is presented first and the ingroup comes second (i.e., intergroup context). In Experiment 3 we further manipulated participant's perception of variability between the ingroup. Based on previous research in other context, we expected participants expressing a stronger black sheep effect in a high entitative condition than in the low entitative condition.

In Experiment 3 (N = 703) we observed the predicted black sheep effect and again the protection of the ingroup culture compared to the outgroup culture. More importantly, we replicated the black sheep effect when the outgroup newsflash came first. This is in line with the idea that presenting a threatening outgroup beforehand may elicit harsher judgments of the ingroup deviant. The entitativity manipulation did not succeed in affecting the judgments. It may be challenging increasing or decreasing entitativity for social groups such as Germans. And furthermore, it may be challenging increasing the black sheep effect and the culture protection in such a deviant context (i.e., ceiling effects). All in all, this fourth contribution provides interesting insights to the existing literature. In contrast to prevalent negative attitudes towards Muslims, we observe a black sheep effect, which is the harsher condemnation of the ingroup perpetrator than the outgroup perpetrator. At cultural level the exoneration of the ingroup may hint to the maintenance of positivity of the whole ingroup. Based on the order effects (presentation order of the newsflashes) which so far has not been discussed as part of the black sheep effect literature, our data shows that it matters whether participants are in an intragroup or intergroup context and that making a threatening outgroup salient elicits the black sheep effect.

Manuscript 4

"(N)One of us but all of them!"
Ingroup favoritism on individual and group level in the context of deviant behavior

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

Authors' note

[1]University of Marburg, Gutenbergstr. 18, 35037 Marburg, Germany

[2]University catholique de Louvain, Belgium

This research was supported by a grant from the German Research Foundation (Research Training Group 2271: "Breaking Expectations", project number: 290878970, project 11).

We wish to thank Sophie Gosch, Amelie Wilhelm, and Anna Schilp for their support in study preparation. We further wish to thank Ulrich Wagner for the exchange and his suggestions especially as part of Experiment 3.

Correspondence concerning this article should be addressed to Zahra Khosrowtaj. E-mail: zahra.khosrowtaj@uni-marburg.de.

Disclosure statement No potential conflict of interest was reported by the authors.

ORCID IDs

Zahra Khosrowtaj: 0000-0001-5327-9656

Sarah Teige-Mocigemba: 0000-0002-7516-1293

Vincent Yzerbyt: 0000-0003-1185-4733

Abstract

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

Keywords: *black sheep effect, ingroup protection, outgroup derogation, intergroup attribution, culture blaming, entitativity, discrimination*

"(…) 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." (Karim 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 (Schmalz, 2023; Tageschau, 2023; Tagesschau, 2023). Among those, 45 were of German nationality, followed by 27 Afghans and 21 from Syria (Schmalz, 2023; Tagesschau, 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 endeavor explores the viability of two strategies to protect the ingroup, namely perceiving an ingroup deviant as one bad apple while in face of an outgroup deviant his culture is at stake. We examined these two strategies in the context of intra- and intergroup situations that involves Germans as ingroups and Muslims[11] as outgroup.

Black Sheep Effect and culture blame

Previous research has shown that we tend to more extremely blame deviant ingroup than comparable outgroup members for maintaining the positivity of the whole ingroup (Abrams et al., 2000, 2002, 2003; Marques, 1990; Marques et al., 1988; Marques & Yzerbyt, 1988; Yzerbyt et al., 2000). For instance, poor speeches by ingroup members were judged more negatively than poor speeches by outgroup members (Marques & Yzerbyt, 1988). This has been conceptualized as a sophisticated form of ingroup favoritism as the

[11] 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).

deviant member is no longer perceived as an ingroup member (Marques et al., 1988; Marques & Yzerbyt, 1988).

A recent work hints to ingroup protection on individual and on cultural level (Khosrowtaj, Yzerbyt, et al., 2023). Specifically, 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 while culturally, the ingroup was evaluated more leniently than the outgroup culture (Khosrowtaj, Yzerbyt, et al., 2023).

Interestingly enough, this pattern is reminiscent of the intergroup attribution bias and the ultimate attribution error. A negative behavior 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 behavior 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 behavior perpetrated by a black (white) person was attributed to personal (situational) factors (Duncan, 1976). Building on this line of work, the present series of experiments replicates and extends Khosrowtaj, Yzerbyt, et al.'s (2022) findings. Our aim is to shed further light on people's comparative judgment 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. Specifically, we predicted ingroup favoritism in form of the black sheep effect at the individual level and leniency towards the ingroup on cultural level. We did so by building on the current debates on the societal threats posed by Muslims and the Muslim culture in Western countries.

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 the 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 behavior towards Muslim appearing targets as part of a shooter paradigm (Unkelbach et al., 2008) or even the discrimination in hiring contexts (Stasio et al., 2019; Unkelbach et al., 2010). Relying on the integrated threat theory (Stephan & Stephan, 1985, 2000), different types of threat, with safety threat standing high on the list, were identified as

response to refugee immigration. This clearly results in negative attitudes towards refugees and support of migration restriction (Landmann et al., 2019).

Media analyses also reveal the presence of a representational bias of 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, Biermann, et al., 2022). Specifically, participants imagined reading a newspaper article depicting a crime before indicating their interest for a list of (non)-stereotypic information categories that they wished to know more about. In line with the representational bias, participants indicated higher interest for 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, Biermann, et al., 2023). Along with other findings (e.g., Fitzgerald et al., 2012; Hirtenlehner, 2019), data such as these illustrate that, in Western countries, Muslims or Muslim-appearing targets (Stasio et al., 2019; Unkelbach et al., 2008, 2010) elicit threat and are associated with violence and crimes. Given our present aim to investigate ingroup protection on individual and group level, we decided to rely on this specific societal context.

The present experiments

The black sheep effect has so far been investigated at the level of the individual member (Marques et al., 1988; Marques & Yzerbyt, 1988). In the larger context of the biased portrayals of Muslim perpetrators (Kearns et al., 2019), the present experiments build upon the observation of direct attributions of physical violence to an alleged Islamic culture (Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022).

Specifically, we conducted three experiments with the aim 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 cultural level, that is, less ingroup culture blaming than outgroup culture blaming (Khosrowtaj, Yzerbyt, et al., 2023). In all experiments, we used alleged newspaper articles (Exp. 1) or newsflashes (Exp. 2-3) where a perpetrator is accused of rape. We manipulated group membership using different names (cf. Kauff, 2022, for a similiar approach using names for categorization) and nationalities (cf. Khosrowtaj, Yzerbyt, et al., 2023). We then assessed participants judgments on individual and cultural level. We predicted a black sheep effect on individual

blame measures (detailed below) while on cultural level, we expected to observe a protection of the ingroup culture.

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 judgments at both individual and cultural level. We pre-registered all the projects reported here (see for Exp. 1: https://osf.io/rqp7c, for Exp. 2: https://osf.io/t8f64, for Exp. 3: https://osf.io/vs8qz). The data and analysis syntax for all three experiments are available at OSF: https://osf.io/3mgj5/?view_only=1a3d3184941b46598ae81d4feb48d2bf.

Experiment 1

The goal of Experiment 1 was to test ingroup protection at the individual and at the cultural level. In line with a host of previous demonstrations of the black sheep effect, we decided to present participants with an intragroup couple. Specifically, 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 allows examining judgments 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 of 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' judgments and behavior. In line with these data, we wanted to examine the impact of mere outgroup salience (cf. Yuki & Yokota, 2009) by means of a priming manipulation on the judgments about ingroup and outgroup perpetrators as well as their cultures.

In sum, Experiment 1 adopted a 2 (couple: ingroup vs. outgroup) x 2 (priming manipulation: outgroup salient vs. outgroup not salient) factorial design. We expected 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 behavior more as rape in case of an ingroup than an outgroup perpetrator. As for the cultural level, we further expected 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. We also expected outgroup salience to influence the classification of the perpetrator's behavior 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.

Methods

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, $\alpha = 0.05$, Power = 0.95 and an estimated small effect size $f^2 = .05$. This analysis resulted in a total sample size of $N = 444$. To take into account possible dropouts, we aimed for a sample size of $N = 500$.

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 the purpose of 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, mother tongue) were assessed at T1.

As pre-registered, we examined the time spent between the priming manipulation and the second part of the study and excluded six participants with a dwell time +3 SD above the mean ($M = 14.43$ ms, $SD = 24.90$). Thus, the following analyses rely on a total sample of $N = 437$ ($n = 160$ female, $n = 224$ male, $n = 2$ diverse, $n = 51$ did not indicate their

gender) with an average age of $M = 31.16$ [12]years ($SD = 10.41$; range: 18 – 68 years; $n = 6$ participants did not indicate their age).

Procedure and measures

Overall procedure. Participants worked through two alleged separate studies. The first part (for participants, the first study of T2) involved the priming manipulation (outgroup salient vs. outgroup not salient) and the second part (for participants the second study of T2) included a fictitious newspaper article (see below) and the assessment of blame and verdict judgments (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 themselves had been victim of sexual violence (sexual abuse $n = 38$, forced intercourse, $n = 24$, almost forced to engage in sexual acts, $n = 40$)[13]. 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, five of the ten questions made the outgroup salient (i.e., "How many Muslims live in Germany?", "How many refugees live in Germany?").

[12] Note that one participant indicated an age of '-1'. As the response patterns did not reveal any suspicion of concern, we did not exclude this subject from the whole analysis. As participants in prolific have to be at least 18 years old, we only excluded this subject for providing the correct descriptive according to age, but all the reported statistical analyses include this mentioned participant.

[13] Leaving out all participants with 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.

Newspaper article. We adapted the fictious newspaper article from previous work (Khosrowtaj, Yzerbyt, et al., 2023), but only included couples with the same ethnic background. The newspaper articles 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.

Judgments. 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 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?" and "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 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 behavior 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, Yzerbyt, et al., 2023), 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 behavior 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).

Results

To test our predictions, we ran a series of multiple regression models. We used 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.

Individual blame

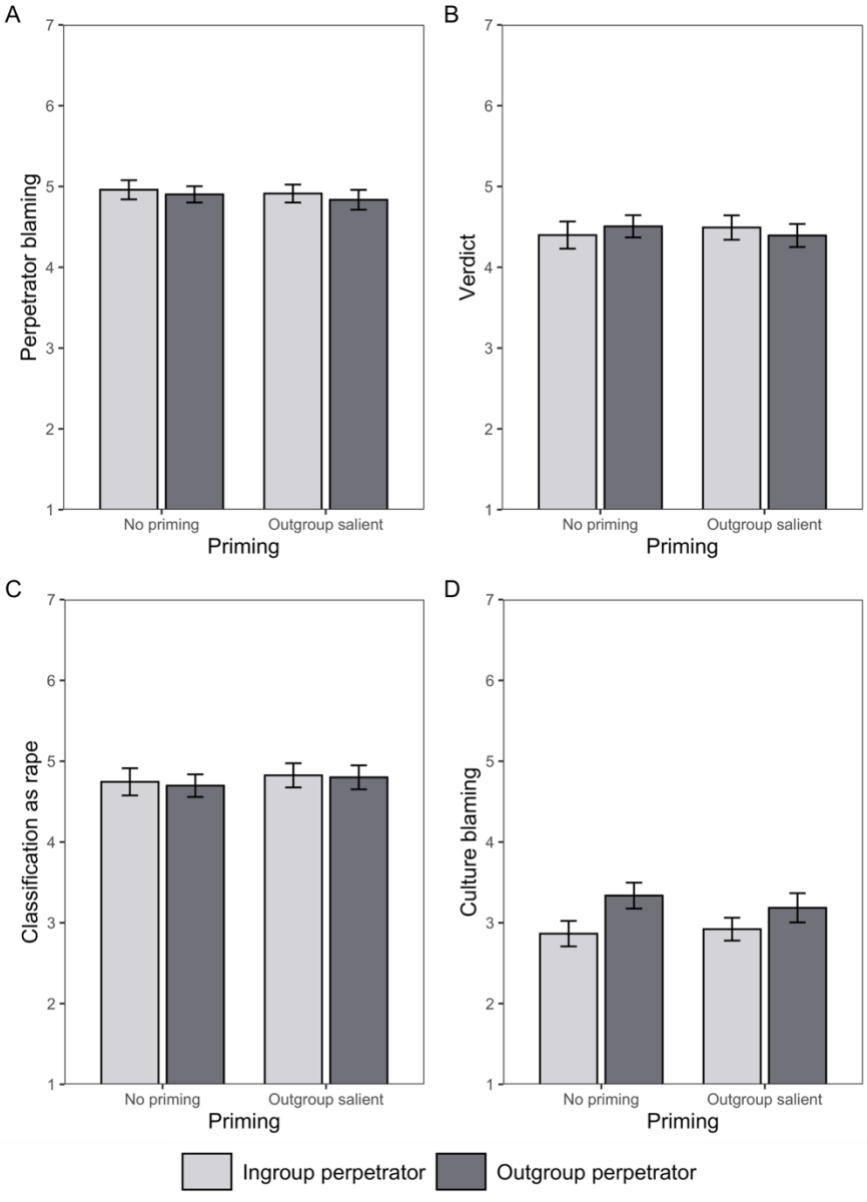
Perpetrator blaming. The analysis with perpetrator blaming as criterion revealed no significant main or interaction effects, all p's ≥ .55 (see Figure 1, panel B).

Verdict. The analysis with verdict as criterion showed no significant main or interaction effect, all p's ≥ .49 (see Figure 1, panel A).

Classification as rape. The analysis with classification of the incident as rape revealed no significant effect of couple nor priming manipulation, all p's ≥ .55 (see Figure 1, panel B).

Culture blaming. The analysis with perpetrator's culture blaming as criterion indicated the predicted main effect of couple, b = .37, $t(426)$ = 2.29, p = .023, η_p^2 = .012, 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's ≥ .520 (see Figure 1, panel C).

Figure 1: Judgments 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 standard error of the mean.

Discussion

In contrast to previous work (Khosrowtaj, Yzerbyt, et al., 2023), we did not find a black sheep effect, whether on perpetrator blame, verdict, or classification of the perpetrator's behavior 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 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 of an outgroup member is more likely to emerge (Otten & Gordijn, 2014; van Prooijen, 2006, but see Khosrowtaj et al., 2022). 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 culture judgments. In all likelihood, the five outgroup questions that were embedded in an alleged quiz did not prove sufficient to make the outgroup salient. Put differently, the priming manipulation included in the alleged first study did not 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 perpetrator and culture judgments.

Considering these limitations, we decided to conduct another experiment to test ingroup protection at the individual and at 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.

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-subjects design), we expected a black sheep effect to occur with harsher judgments of the ingroup than of the outgroup perpetrator on all measures targeting the individual: perpetrator blaming, verdict, and classification of perpetrator's behavior as rape. We further expected to replicate the exoneration of the ingroup culture than the outgroup culture. Finally, we also aimed to examine whether the order of presentation affects the judgments. If any, order effects may arise due to the outgroup first condition where the intergroup context is salient (e.g., Haslam & Oakes, 1995).

Sample size estimation

We ran a power analysis using 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 perpetrator. Power analysis was based on the main effect of factor couple in Experiment 1 for culture blaming. Here, we estimated a d = .24. This resulted in N = 300 for achieving a Power of 94.9%.

Participants

A total of N = 303 participants from the Prolific Academic (www.prolific.co) participated in the online experiment. We excluded fifteen participants due to our a priori

exclusion criteria. Specifically, we excluded three participants with too short dwell times (< 5 seconds) on the pages showing the vignettes, seven participants with parents with a mother tongue from a predominantly Muslim country, and one participant who participated in a previous similar experiment. Based on recommendations by Judd et al. (2011), we inspected the studentized deleted residuals for the dependent variables as means of the group factor and we excluded nine participants with absolute values of the studentized deleted residuals > +/- 4 (Judd et al., 2011)[14]. Our final analyses relied on a total sample of N = 283 (n = 121 female, n = 154 male, n = 4 diverse, n = 4 did not indicate their gender) with an average age of M = 28.41 years (SD = 7.61; range: 18 – 66 years).[15]

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 judgment 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.

[14] Note that we excluded no participant of Exp. 1 based on this criterion.

[15] 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.

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, Yzerbyt, et al., 2023) and for ruling out any other inter- or intragroup context beyond the one created by the within-subjects design.

Judgments. 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.

Results

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

Individual blame

Perpetrator blaming. The analysis with perpetrator blaming as criterion revealed a significant main effect of perpetrator, $F(1,281) = 5.31$, $p = .022$, $\eta_p^2 = .019$, indicating that participants blamed the ingroup perpetrator more strongly ($M = 6.28$, $SD = .69$) than his outgroup counterpart ($M = 6.21$, $SD = .73$). We further observed a marginally significant effect of order, $F(1,281) = 3.69$, $p = .056$, $\eta_p^2 = .013$, such that the overall mean ratings

were harsher when the ingroup was presented first (M = 6.32, SD = .06) than when the outgroup was presented first (M = 6.17, SD = .06). We further observed a significant interaction, $F(1,281) = 39.35$, $p < .001$, $\eta_p^2 = .123$ (see Figure 2, panel A). To probe this interaction, we ran post-hoc paired samples *t*-tests separately for each order. When the outgroup newsflash was presented first, the ingroup perpetrator was blamed more (M = 6.30, SD = .73) than the outgroup perpetrator (M = 6.04, SD = .79), $t(138) = 6.14$, $p < .001$. When the ingroup newsflash came first, the outgroup perpetrator was blamed more (M = 6.38, SD = .63) compared to his ingroup counterpart (M = 6.26, SD = .65), $t(143) = -2.78$, $p = .006$.[16]

Verdict. The analysis with verdict as criterion again revealed a significant main effect of perpetrator, $F(1,281) = 4.32$, $p = .039$, $\eta_p^2 = .015$, indicating that participants blamed the ingroup perpetrator more strongly (M = 6.28, SD = .82) than his outgroup counterpart (M = 6.21, SD = .88). We further observed a significant main effect of order $F(1,281) = 4.99$, $p = .026$, $\eta_p^2 = .017$, such that the overall mean ratings were harsher when the ingroup was presented first (M = 6.35, SD = .07) than when the outgroup was presented first (M = 6.14, SD = .07). We further observed a significant interaction, $F(1,281) = 63.83$, $p < .001$, $\eta_p^2 = .185$ (see Figure 2, panel B). We again ran post-hoc paired samples *t*-tests, separately for each order. When the outgroup newsflash was presented first, the ingroup perpetrator was blamed more (M = 6.30, SD = .81) than the outgroup perpetrator (M = 5.96, SD = .96), $t(138) = 6.81$, $p < .001$. When the ingroup newsflash was presented first, the outgroup perpetrator was blamed more (M = 6.45, SD = .73) than his ingroup counterpart (M = 6.25, SD = .85), $t(143) = -4.34$, $p = .006$.[17]

Classification as rape. The analysis on the classification of the incident as rape revealed a significant perpetrator × order interaction, $F(1,281) = 21.59$, $p < .001$, $\eta_p^2 = .071$ (see Figure

[16] 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 perpetrator ($p = .045$), no effect of order ($p = .23$), and a significant interaction ($p < .001$).

[17] 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 perpetrator ($p = .010$), a significant main effect of order ($p = .017$), and a significant interaction ($p < .001$).

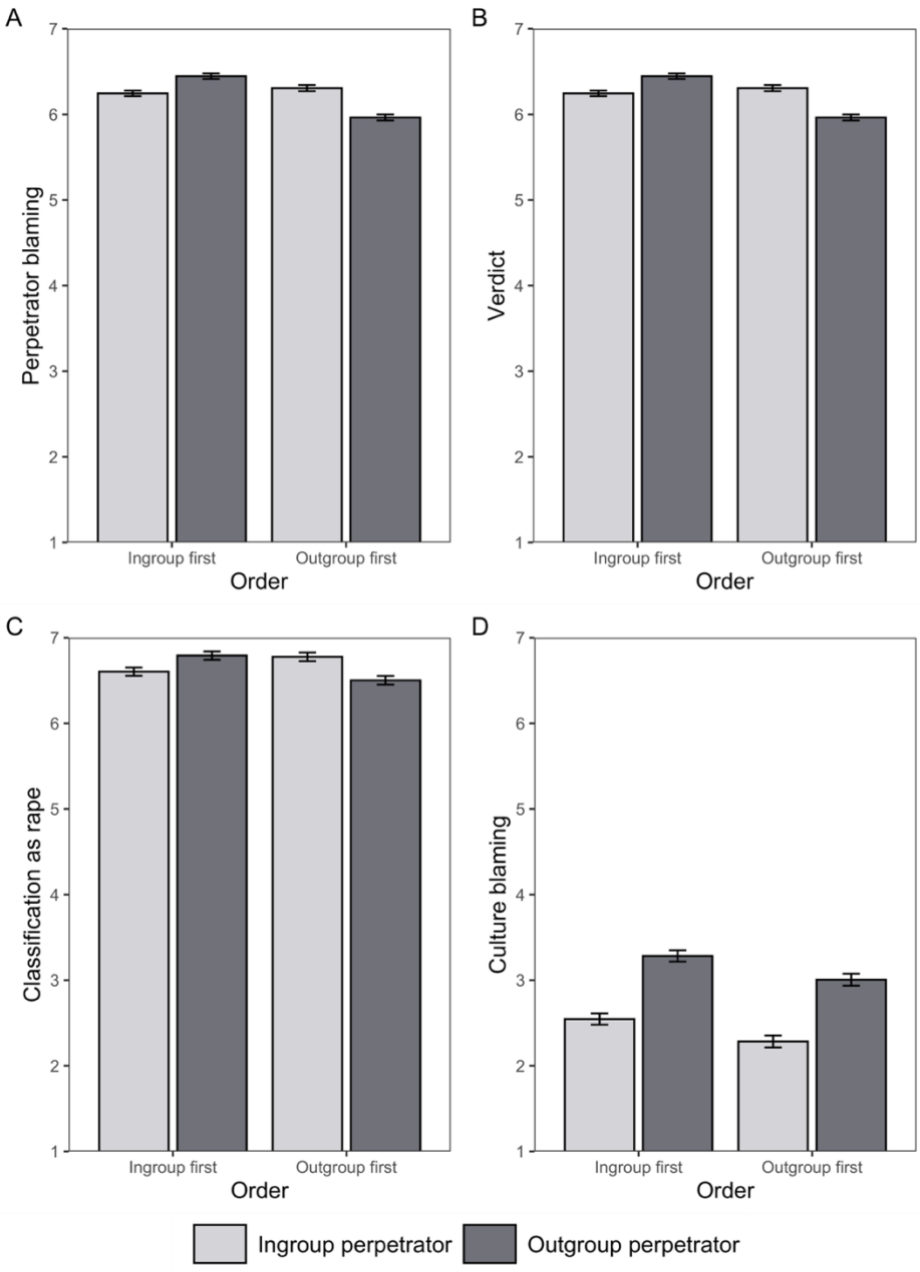
2, panel C). There was no other significant effect, all p' s > .387. To probe this interaction, we ran post-hoc paired samples *t*-tests for each order. When the outgroup newsflash was presented first, the ingroup perpetrator's behaviour was more likely classified as rape (M = 6.77, SD = .55) than the outgroup perpetrator's behaviour (M = 6.50, SD = .84), $t(138)$ = 3.83, $p < .001$. When the ingroup newsflash was presented first, the outgroup perpetrator's behaviour was more likely classified as rape (M = 6.79, SD = .58) than the behaviour of his ingroup counterpart (M = 6.60, SD = .83), $t(143)$ = -2.72, p = .007.[18]

Culture blaming. Turning to culture blaming, the predicted main effect of perpetrator proved significant, $F(1,281)$ = 116.22, $p < .001$, η_p^2 = .293, such that the ingroup culture came across as less responsible of the deviant act (M = 2.42, SD = 1.23) than the outgroup culture (M = 3.15, SD = 1.58). There was no other effect, all p's ≥ . 087 (see Figure 2, panel D).[19]

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

[19] 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 perpetrator ($p < .001$) but no main effect of order (p = .187) nor an interaction effect (p = .481).

Figure 2: Judgments as a function of perpetrator's group membership and order of presentation



Notes. A: Perpetrator blaming as a function of 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 perpetrator's group membership and order of presentation. D: Culture blaming as a function of perpetrator's group membership and order of presentation. Higher scores indicate harsher judgments. Error bars indicate standard error of the mean.

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 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 judgments 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 (Haslam et al., 1995). Indeed, when participants who began with the ingroup perpetrator continued with the second newsflash, they encountered an intergroup context. In all likelihood, they then opted to distance themselves from the outgroup perpetrator (as representing a homogenous threatening group) and expressed harsher judgment 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 judgments given the threat to the ingroup's image (Lewis & Sherman, 2010; Yzerbyt et al., 2000). To test this conjecture, Experiment 3 examined the impact of a manipulation of ingroup entitativity on ingroup protection at the individual and at the cultural level.

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 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 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 entitativity of a national group such as Germans to investigate its influence on judgments of highly deviant behavior. 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 judgment 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 with regard to 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 effects of our experimental entitativity manipulation (i.e., ceiling effects).

Methods

Sample size estimation

We ran a power Analysis using 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 perpetrator.

In Experiment 2, we observed a significant black sheep effect with d_z = .41. We assumed a smaller effect for Experiment 3 of the within-subjects factor perpetrator (reflecting the black sheep effect) due to the entitativity manipulation. Based on our predictions, entitativity shall increase the black sheep effect in the high entitativity condition and lead to a decrease or no black sheep effect in the low entitativity condition. Thus, being more conservative, we assumed a smaller effect size, namely d_z = .20 for both the black sheep effect (reflected in a main effect of perpetrator) and the interaction between perpetrator × entitativity. Using this effect size (.20) revealed a power of .898 with a total of N = 700 participants (n = 175 per condition, i.e., order × entitativity).

Participants and design

A total of N = 750 participants from the Prolific Academic with German as first language participated in the study in two waves. The second wave served for reaching the final sample size given that we had to exclude participants with parents with critical mother tongues (see below) or who failed the manipulation check. In the following, the exclusion criteria are depicted, taking into account both data collections which aimed at reaching our preregistered sample size of 700 subjects.

We excluded participants who (a) had parents with mother tongues from a predominantly Muslim country (n = 21), (b) who failed the manipulation check (n = 14), (c) spent <= 5 seconds on the pages with the vignettes (n = 5, one of them also failing the manipulation check) and (d) who participated in similar previous studies (n = 2). We further observed the studentized deleted residuals for the dependent variables and excluded six participants with absolute values > 4 (Judd et al., 2011). This led to an analysis sample of N = 703. Due to a programming error, participants in the condition of *ingroup* perpetrator as first newsflash in the *context* of cinema did not see the dependent measures regarding verdict items. That is why we report perpetrator blaming only as dependent measure.

Procedure and measures

Overall procedure. We invited participants to take part in two studies. Participants were randomly assigned to either of the entitativity conditions (high vs. low) which marked the

first study for the participants. 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 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.

Subsequently, participants started with the alleged second study, which mimicked Experiment 2. Here the procedure was the same as in Experiment 2. After working through the first newsflash and its blame judgments, participants worked through the second newsflash and the blame items and finished 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 so 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?" (Varying from 1 = not at all similar to 7 = very similar).

Newsflashes. We used the newsflashes from Experiment 2 and adapted the last sentence of each newsflash for leaving room for guilt being slightly more ambiguous as mentioned above. 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.*

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

Results

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

Individual blame

Perpetrator blaming. The first analysis with perpetrator blaming[20] as dependent variable revealed a significant main effect of perpetrator, $F(1,699) = 57.45$, $p < .001$, $\eta_p^2 = .076$, indicating that participants blamed the ingroup perpetrator more strongly ($M = 6.15$, $SD = .78$) than his outgroup counterpart ($M = 5.95$, $SD = .92$). We further observed a significant main effect of order $F(1,699) = 10.62$, $p = .001$, $\eta_p^2 = .015$, such that the overall judgments of the perpetrators were harsher when the ingroup was presented first ($M = 6.14$, $SD = .04$) compared to the case when the outgroup was presented first ($M = 5.96$, $SD = .04$). In addition, we observed a significant interaction of perpetrator × order, $F(1,699) = 99.09$, $p < .001$, $\eta_p^2 = .124$. To probe the interaction, we ran post-hoc paired samples *t*-tests for each order. When the outgroup newsflash came first, the ingroup perpetrator was blamed more harshly ($M = 6.18$, $SD = .76$) than the outgroup perpetrator ($M = 5.73$, $SD = .97$), $t(355) = 11.20$, $p < .001$, $d_z = .594$. When the ingroup newsflash came first, there was no significant difference between the blame attributions except for a marginal trend: the outgroup perpetrator was marginally blamed more harshly ($M = 6.17$, $SD = .81$) compared to his ingroup counterpart ($M = 6.11$, $SD = .80$), $t(346) = -1.97$, $p = .050$, $d_z = .106$. We did not observe any significant main or interaction effect regarding entitativity, see Figure 3, panel A.[21]

---

[20] As the assumption of homogeneity of variances was violated for the repeated measures variable (perpetrator blaming outgroup $p = .001$) we 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 ($p < .001$).

[21] Due to a programming error, participants in the ingroup first and newsflash depicting the cinema context did not work through the verdict items. As such this measure includes missing data in one of four conditions and the results should be interpreted with caution. This measure mirrored the perpetrator blaming patterns with a main effect of perpetrator ($F(1,538) = 14.06$, $p < .001$, $\eta_p^2 = .025$, ingroup perpetrator ($M = 6.21$, $SD = .99$) > outgroup perpetrator ($M = 5.95$, $SD = 1.14$)), a trend for a main effect for order ($F(1,538) = 3.76$, $p = .053$, $\eta_p^2 = .007$, ingroup first ($M = 6.19$, $SD = .07$), outgroup first ($M = 6.03$, $SD = .05$)), and an interaction effect between perpetrator and order, $F(1,538) = 53.72$, $p < .001$, $\eta_p^2 = .091$. For probing the interaction, we ran post-hoc paired samples

However, it is noteworthy, that the manipulation check indicated that the entitativity manipulation was successful as it affected the degree of similarity and connectedness perceived by participants, $F(1, 701) = 87.39$, $p < .001$, $\eta_p^2 = .111$. 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 on the classification of the perpetrator's behaviour as rape[22] revealed a significant effect of perpetrator, $F(1,699) = 49.22$, $p < .001$, $\eta_p^2 = .066$, that is the ingroup perpetrator's behaviour was classified more likely as rape ($M = 6.54$, $SD = .89$) than the outgroup perpetrator's behaviour ($M = 6.21$, $SD = 1.15$). We further observed a significant effect of order, $F(1,699) = 5.98$, $p = .015$, $\eta_p^2 = .008$: again, the overall judgments for the classification as rape were harsher when the ingroup newsflash came first ($M = 6.45$, $SD = .04$) than when the outgroup newsflash came first ($M = 6.31$, $SD = .04$). In addition, we observed a significant interaction between perpetrator × order, $F(1,699) = 26.45$, $p < .001$, $\eta_p^2 = .036$. Post-hoc paired samples *t*-tests for each order revealed that when the outgroup newsflash came first, the ingroup perpetrator's behaviour was classified more as rape ($M = 6.59$, $SD = .79$) than the outgroup perpetrator's behaviour ($M = 6.03$, $SD = 1.25$), $t(355) = 7.96$, $p < .001$, $d_z = .422$. When the ingroup newsflash came first, there was no significant difference between the classification as rape ($M = 6.49$, $SD = .98$; $M = 6.40$, $SD = .99$, for ingroup and outgroup perpetrator respectively), $t(346) = 1.46$,

---

t-tests separately for the levels of factor order. When the outgroup vignette was presented first, the ingroup perpetrator was blamed more harshly ($M = 6.28$, $SD = .87$) compared to the outgroup perpetrator ($M = 5.79$, $SD = 1.21$), $t(355) = 9.38$, $p < .001$, $dz = .497$. When the ingroup vignette was presented first, the outgroup perpetrator was blamed more harshly ($M = 6.26$, $SD = .93$) compared to the outgroup perpetrator ($M = 6.09$, $SD = 1.18$), $t(185) = -2.40$, $p = .017$, $|dz| = .176$. We did not observe any significant main or interaction effect regarding entitativity (entitativity, $F(1,538) = 1.32$, $p = .251$; perpetrator × entitativity, $F(1,538) = .02$, $p = .880$; perpetrator × order × entitativity, $F(1,538) = 1.06$, $p = .303$).

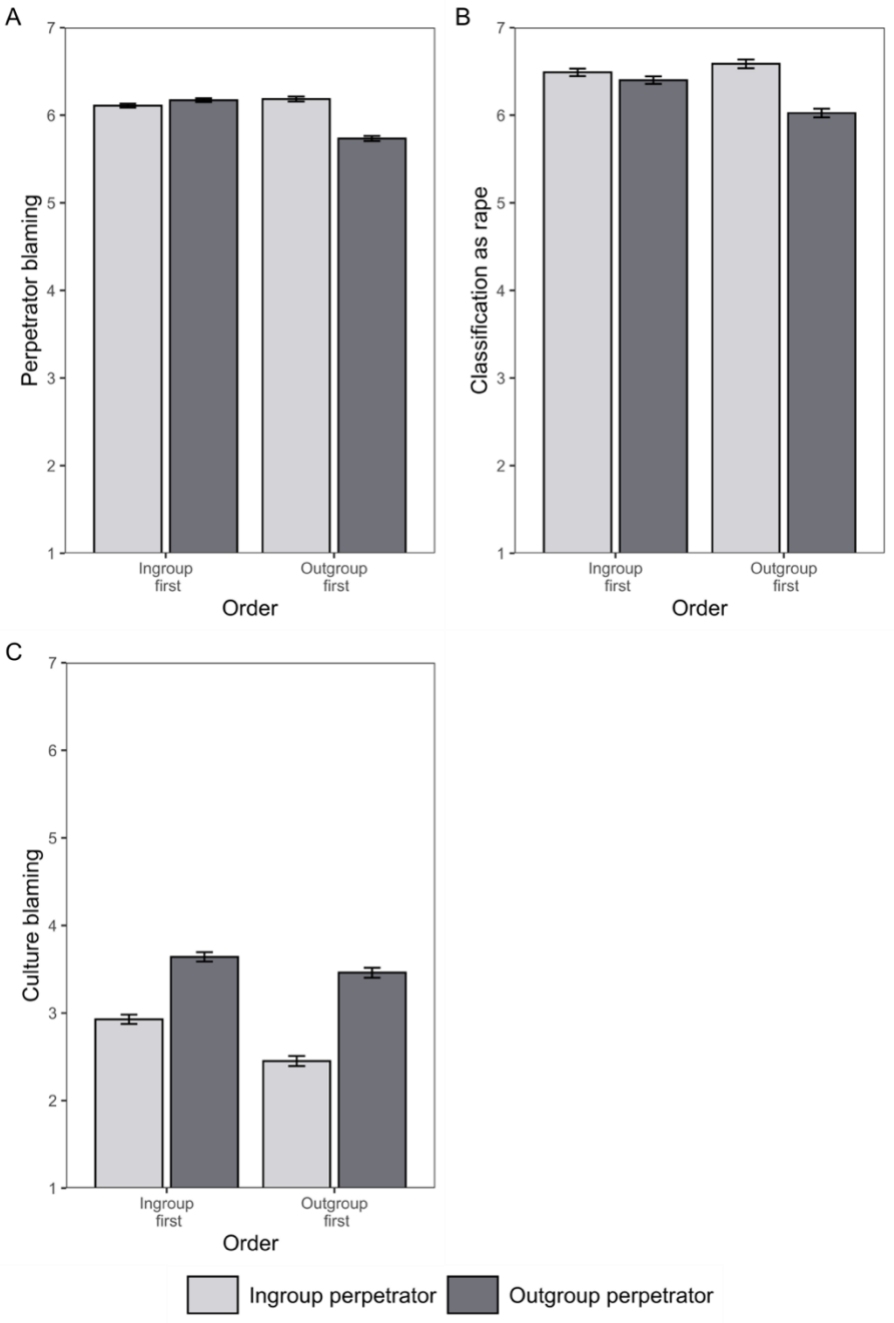
[22] As the assumption of homogeneity of variances was violated for the repeated measures variable (classification as rape ingroup $p = .001$, outgroup $p < .001$) we conducted a robust ANOVA which was in line with the observed significant results (perpetrator: $p < .001$, order: $p = .009$) interaction effect ($p < .001$).

p = .146 (see Figure 3, panel B). We did not observe any significant main or interaction effect regarding entitativity.

Culture blaming. We again replicated the predicted effect of perpetrator on culture blaming, $F(1,699) = 242.05$, $p < .001$, $\eta_p^2 = .257$, such that the ingroup culture was judged more leniently (M = 2.68, SD = 1.47) than the outgroup culture (M = 3.55, SD = 1.68). In contrast to Experiment 2, we also observed an effect of order, $F(1,699) = 8.97$, $p = .003$, $\eta_p^2 = .013$, such that the overall judgments were harsher when the ingroup newsflash came first (M = 3.28, SD = .08; M = 2.96, SD = .07, for ingroup and outgroup newsflash, respectively). In addition, we observed a significant interaction[23] between perpetrator × order, $F(1,699) = 6.83$, $p = .009$, $\eta_p^2 = .010$ (see Figure 3, panel C). Paired samples t-tests for each order 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), $t(355) = -12.44$, $p < .001$, $|d_z| = .659$. In a similar vein, but less pronounced, 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), $t(346) = -9.45$, $p < .001$, $|d_z| = .509$. We did not observe any significant main or interaction effect regarding entitativity.

[23] As the assumption of homogeneity of variances was violated for the repeated measures variable (culture blaming ingroup p = .006) we conducted a robust ANOVA on trimmed means. This indicated both significant main effects (perpetrator: $p < .001$, order: $p = .003$) but no interaction effect (p = .106).

Figure 3: Judgments as a function of perpetrator's group membership and order of presentation



Notes. A: Perpetrator blaming as a function of perpetrator's group membership and order of presentation, B: Classification as rape as a function of perpetrator's group membership and order of presentation. C: Culture blaming as a function of perpetrator's group membership and order of presentation. Higher scores indicate harsher judgments. Error bars indicate standard error of the mean.

Discussion

With regard to our replication aim, 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 outgroup perpetrator. They further 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 judgments towards the ingroup perpetrator when they had first read the outgroup newsflash. This pattern was present also on the classification of the perpetrator's behaviour as rape: 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 judgments 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 judgments when the outgroup newsflash was presented first. Again, this confirms that making an intergroup context salient reinforces the protection of the ingroup (e.g., 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. Further, 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 judgments.

General discussion

In the present work, we investigated ingroup protection at the individual and cultural level. We built on recent work showing harsher judgments of individual ingroup

perpetrators (i.e., black sheep effect) on the one hand and more lenient judgment of the ingroup as a whole on the other hand (Khosrowtaj, Yzerbyt, et al., 2023). In Experiment 1, we tested whether the mere salience of a threatening outgroup may elicit harsher judgments 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 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.

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 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 a protection of the ingroup both on individual and cultural level. 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. These results may also hint to a derogation of the outgroup as a whole. Both judgments on individual and cultural level are comparative, that is one may take the perspective that one outgroup perpetrator is not judged more harshly than his ingroup counterpart, while on 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 in fact 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 it was replicated in Experiment 3. This pattern suggests that the presence of an intergroup context constitutes a necessary condition for the black sheep effect to occur.

This difference between the ascription of blame at the individual and at cultural level hints to a possible proximity to the concept of shifting standards (e.g., Biernat & Manis, 1994). Different judgment standards may have been used for in- and outgroup deviants

based on group stereotypes (Biernat et al., 1991). Even though the focus of the present work did not rely on the use of different dependent variables (i.e., subjective vs. objective items), the black sheep effect may be seen as a form of more demanding standard for the ingroup perpetrator and more favourable judgment (leniency) towards the outgroup perpetrator based on the stereotype that "Muslims/Afghans are violent" (Biernat et al., 1991; Linville & Jones, 1980). Shifting standards have been observed to be motivated by ingroup protection (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 et al. (2010) targeted only the ingroup. As such, the present contribution extends the examination of different standards as we compared judgments attributed to both in- and outgroup perpetrators on individual and cultural level.

Limitations and future directions

Regarding the failure of our priming manipulation, we suggest that the operationalization was 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 has been discussed alongside the order effects of Exp. 2-3. Along similar lines, our entitativity manipulation of Experiment 3 did not affect participants' judgments 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 taking into account 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 on culture blaming arise as a cause 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.

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 background, and refugees.

Conclusions

The present work offers a number of interesting insights to 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 more harshly judge a German perpetrator compared to his Afghan counterpart. This black sheep effect may seem paradoxical but it 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 in an intergroup context. Making a threatening outgroup salient facilitates the emergences of the black sheep effect as a way to express that such a deviant behavior is not typical for the ingroup (Marques & Yzerbyt, 1988; Yzerbyt et al., 2000).

Third, the ingroup is protected on cultural level. The comparative leniency towards the ingroup culture may be a form of outgroup discrimination on cultural level which may incriminate 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 while in case of an outgroup perpetrator, it is not him but his culture who is derogated. These data suggest that the outgroup deviant is not so much treated as an individual but rather as an instance of an otherwise homogenous and indeed negative outgroup. Raising the level of awareness regarding this judgmental strategy may be crucial means for decreasing hostility against Muslims and reducing anti-Muslim prejudices (Bruneau et al., 2020; Gallardo et al., 2021). Investigating this aspect is undoubtedly an avenue for future research.

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), Article 5. https://doi.org/10.1037/0022-3514.78.5.906

Abrams, D., Rutland, A., Cameron, L., & Marques, Josém. (2003). The development of subjective group dynamics: When in-group bias gets specific. *British Journal of Developmental Psychology*, *21*(2), 155–176. https://doi.org/10.1348/026151003765264020

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

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

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

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*, 13684302211021664. 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

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]. Berlin: 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

Kauff, M. (2022). Muslim = Terrorist? Attribution of violent crimes to terrorism or mental health problems depend on perpetrators' religious background. *The Journal of Social Psychology*, *0*(0), Article 0. 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. (2023). "He must be a foreigner" On biased information search towards in- and outgroup wrongdoers. *Manuscript Submitted*.

Khosrowtaj, Z., Yzerbyt, V., Süssenbach, P., & Teige-Mocigemba, S. (2023). On the effect of perpetrator's description on inter- and intragroup guilt attributions. *Manuscript Submitted*.

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

Landmann, H., Gaschler, R., & Rohmann, A. (2019). What is threatening about refugees? Identifying different types of threat and their association with emotional responses and attitudes towards refugee migration. *European Journal of Social Psychology*, *49*(7), Article 7. https://doi.org/10.1002/ejsp.2593

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 Subtypes for Ingroup Members. *Personality and Social Psychology Bulletin, 39*(5), Article 5. https://doi.org/10.1177/0146167213479978

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

Schmalz, M. W., AFP,Alexander. (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

Seelig, L. (2023, January 13). Karim Fereidooni: 'Eltern haben Angst um den Bildungserfolg ihres kleinen Maximilian'. *Die Zeit*. 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

Stasio, V. D., 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*, *0*(0), Article 0. https://doi.org/10.1080/1369183X.2019.1622826

Stephan, W. G., & Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, *41*(3), Article 3. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In *Reducing prejudice and discrimination* (pp. 23–45). Lawrence Erlbaum Associates Publishers.

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Muslim 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

Tageschau. (2023, 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. (2023, 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 violence: '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

Yuki, M., & Yokota, K. (2009). The primal warrior: Outgroup threat priming 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, Forum Berlin.

Supplementary materials

Experiment 3: Entitativity manipulation

<u>High entitativity German wording as in the Experiment:</u>

Entnommen und gekürzt aus: Fischer, J. & Vikas, R. (2020). How we perceive ourselves: results of a representative survey in Germany. Journal of Cultural Psychology, 1(3), 257-263.

Insgesamt deuten die Ergebnisse der bevölkerungsrepräsentativen Studie darauf hin, dass Deutsche besonders im Vergleich zu anderen europäischen Ländern als ähnlich und zusammengehörig beschrieben werden können.

Auf die Frage, welche Werte den Teilnehmenden im Leben wichtig und erstrebenswert erscheinen, haben nicht weniger als 78% der Deutschen dieselben Top 3 Werte ausgewählt: körperliches und psychisches Wohlbefinden, Freiheit und enge Beziehungen zu anderen Menschen.

Ein ähnliches Muster ergab sich bei den Persönlichkeitsmerkmalen, die die Teilnehmenden als für sich typisch ansehen. Hier wählten sogar fast 83% der Befragten vergleichbare Merkmale auf den ersten fünf Plätzen.

Interessanterweise scheinen sich die Deutschen nicht nur in ihren Werten und Persönlichkeitsmerkmalen sehr ähnlich zu sein, auch in Bezug auf alltägliche Gewohnheiten wie beispielsweise Ess- und Trinkverhalten wurden bei mehr als zwei Dritteln der Teilnehmenden sehr hohe Überschneidungen gefunden.

In Bezug auf das Zusammengehörigkeitsgefühl in Deutschland gaben annähernd 73,8% der Befragten an, dass es Ihnen wichtig ist, dass sie sich gegenseitig unterstützen und insbesondere in Krisenzeiten und Not zusammenhalten.

Aus den Ergebnissen lässt sich klar schlussfolgern, dass Deutsche im Vergleich zu anderen europäischen Ländern als zusammengehörig - im Sinne einer Einheit – beschrieben werden können.

English translation:

Taken and abridged from: Fischer, J. & Vikas, R. (2020). How we perceive ourselves: results of a representative survey in Germany. Journal of Cultural Psychology, 1(3), 257-263.

Overall, the results of the population-representative study indicate that Germans can be described as similar and belonging together, especially in comparison to other European countries.

When asked which values the participants considered important and worth striving for in life, no fewer than 78% of Germans chose the same top 3 values: physical and mental well-being, freedom and close relationships with other people.

A similar pattern emerged for the personality traits that the participants considered typical of themselves. Here almost 83% of those questioned chose comparable characteristics in the first five places.

Interestingly, Germans not only seem to be very similar in their values and personality traits, but also in relation to everyday habits such as eating and drinking behavior, very high overlaps were found in more than two thirds of the participants.

With regard to the feeling of togetherness in Germany, almost 73.8% of those surveyed stated that it is important to them that they support each other and stick together, especially in times of crisis and need.

From the results it can be clearly concluded that Germans can be described as belonging together - in the sense of a unit - in comparison to other European countries.

Low entitativity: German wording as in the Experiment:

Entnommen und gekürzt aus: Fischer, J. & Vikas, R. (2020). How we perceive ourselves: results of a representative survey in Germany. Journal of Cultural Psychology, 1(3), 257-263.

Insgesamt deuten die Ergebnisse der bevölkerungsrepräsentativen Studie darauf hin, dass

Deutsche besonders im Vergleich zu anderen europäischen Ländern als einander unähnlich und individuell beschrieben werden können.

Auf die Frage, welche Werte den Teilnehmenden im Leben wichtig und erstrebenswert erscheinen, haben weniger als 22% der Deutschen dieselben Top 3 Werte ausgewählt: körperliches und psychisches Wohlbefinden, Freiheit und enge Beziehungen zu anderen Menschen.

Ein ähnliches Muster ergab sich bei den Persönlichkeitsmerkmalen, die die Teilnehmenden als für sich typisch ansahen. Hier wählten gerade mal 17% der Befragten vergleichbare Merkmale auf den ersten fünf Plätzen.

Interessanterweise scheinen sich die Deutschen nicht nur in ihren Werten und Persönlichkeitsmerkmalen zu unterscheiden, auch in Bezug auf alltägliche Gewohnheiten wie beispielsweise Ess- und Trinkverhalten konnten bei nur 26,2% der Teilnehmenden deutliche Überschneidungen gefunden werden.

In Bezug auf das Zusammengehörigkeitsgefühl in Deutschland gaben 26,2% der Befragten an, dass es ihnen wichtig ist, dass sie sich gegenseitig unterstützen und insbesondere in Krisenzeiten zusammenhalten.

Aus den Ergebnissen lässt sich klar schlussfolgern, dass Deutsche im Vergleich zu anderen europäischen Ländern eher als einander unähnlich und individuell beschrieben werden können.

English translation:

Taken and abridged from: Fischer, J. & Vikas, R. (2020). How we perceive ourselves: results of a representative survey in Germany. Journal of Cultural Psychology, 1(3), 257-263.

Overall, the results of the population-representative study indicate that Germans can be described as dissimilar and individual, especially in comparison to other European countries.

When asked which values the participants considered important and worth striving for in life, less than 22% of Germans selected the same top 3 values: physical and mental well-being, freedom and close relationships with other people.

A similar pattern emerged for the personality traits that the participants considered typical of themselves. Here just 17% of those surveyed chose comparable characteristics in the first five places.

Interestingly, Germans not only differ in their values and personality traits, but also in relation to everyday habits such as eating and drinking behavior, clear overlaps were found in only 26.2% of the participants.

With regard to the feeling of togetherness in Germany, 26.2% of those surveyed stated that it is important to them that they support each other and stick together, especially in times of crisis.

From the results it can be clearly concluded that Germans are more likely to be described as dissimilar and individual compared to other European countries.

General Discussion

As part of this dissertation, we examined the impact of several characteristics of a) the perpetrator, b) the deviant act as well as c) the evaluator on blame attributions in the context of deviant behavior. We used the framework of coping with ingroup deviance model while going beyond the perpetrator by further shedding light on the impact of victim's ethnicity and more importantly by examining attributions on cultural level. The following paragraph summarizes the main findings of the present dissertation.

Summary of the main findings

Based on media analysis hinting to a biased representation of refugees and Muslims and the attribution of deviant acts to their background and religious affiliation (Kearns et al., 2019; Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022) the first contribution of this work tested whether German participants indicate indeed higher interest towards stereotypic information categories (e.g., religious affiliation) in face of a perpetrator from a predominantly Muslim country than his German counterpart. This first empirical test revealed that after imagining a newspaper article about a deviant act, participants indicate higher interest towards stereotypic information categories in case of an outgroup than ingroup perpetrator. Further, participants' judgments on stereotypicality clarified their awareness about the stereotypic information categories. As such the biased representation as part of the media is existent in the information search of participants.

As part of the second manuscript, we tested whether an ingroup perpetrator compared to his outgroup counterpart would be judged more harshly relying on literature on the black sheep effect (Marques & Yzerbyt, 1988). We further aimed to test whether guilt certainty would interact with the perpetrator's ethnicity. Recall the so called benefit of the doubt: it has been shown that when ingroup's guilt is not certain then outgroup discrimination occurs, in case of guilt being certain the black sheep effect is expressed (van Prooijen, 2006). The black sheep effect was only observed on the dependent measures used by van Prooijen (2006), namely anger and hostile affect. We did not observe the interaction between ethnicity and guilt certainty even though the guilt certainty manipulation was successful. Thus, we proceeded with Exp. 2 of manuscript 2, where we tested our idea that perceiving the outgroup as less human may affect the judgments of the deviant behavior leading to unequal blame attributions. As such, we used an experimental manipulation for reducing infrahumanization (increasing the ascription of secondary emotions) based on work conducted by Davies et al. (2018). In contrast to what prior research on

infrahumanization has shown (Leyens et al., 2000, 2001), we observed valence differences as a result of our manipulation. Participants in the infrahumanization condition ascribed higher degrees of positive secondary emotions but surprisingly fewer negative secondary emotions to the outgroup. Recall that based on the infrahumanization theory independent of valence, secondary emotions are preserved for the ingroup and thus the manipulation intended to increase the ascription of both positive and negative secondary emotions to the outgroup. One explanation for this may be due to the infrahumanization manipulation which introduced a positive act of the outgroup (helping the ingroup). The proceeding assessment of the emotions may have been conflicting in case of the negative emotions. Davies et al. (2018) observed a reduction of infrahumanization following an outgroup helping the ingroup manipulation after a natural catastrophe (negative incident) while assessing only negative (primary and secondary) emotions. Previous research did not always include both valences of emotions (Cuddy et al., 2007; Davies et al., 2018; Demoulin et al., 2009), as such it is difficult to examine differential effects.

However, the experimental infrahumanization manipulation did not influence blame judgments: We observed the black sheep effect independent of the infrahumanization manipulation. Interestingly, as part of both experiments of manuscript 2, the victim was blamed more harshly when the perpetrator stemmed from the outgroup. These results pointed to a shift of blame from the outgroup perpetrator to the ingroup perpetrator and the ingroup victim.

In addition, we observed correlational patterns regarding the ascribed emotions. With increasing attribution of secondary positive emotions, participants blamed the outgroup perpetrator less severely. Mirroring this effect, with increasing attribution of secondary positive emotions to the outgroup the victim was blamed more harshly when the perpetrator stemmed from the outgroup. These correlational patterns concerning the positive secondary emotions are also in line with a shift of blame away from the outgroup perpetrator.

We further used the information categories (tested as part of manuscript one) and asked participants about which information categories they wished to be informed further. Participants indicated higher interest towards the information categories religious affiliation, cultural and ethnic background as well as civil status in case of an outgroup than ingroup perpetrator.

For further investigating the shift of blame from the outgroup perpetrator to the ingroup perpetrator, we continued with contributions which were depicted as part of manuscript 3. Again, we tested whether German participants would blame the ingroup

perpetrator more harshly than his outgroup counterpart, replicating the black sheep effect. In addition, we examined whether further descriptions of the perpetrator would matter regarding the judgments. We examined whether participants would exonerate a positively described perpetrator (warm and competent) more readily than his counterpart described in negative terms (cold and incompetent) using the stereotype content model (for a recent review, see Abele et al., 2021; Fiske et al., 2002). Results indicated both the black sheep effect as well as the exoneration of the positively described perpetrator. Again, the victim was judged more harshly when the perpetrator stemmed from the outgroup. However, based on the first experiment of manuscript 3, we could not conclude whether the victim is blamed in an intergroup dyad condition (George & Martínez, 2002) or only when she belonged to the ingroup and the perpetrator to the outgroup.

That is why we manipulated the victim's ethnicity as part of Exp. 2 of manuscript 3. More importantly, in line with the idea of a shift of blame (from the outgroup individual to the ingroup perpetrator and to the ingroup victim), we examined blame attributions towards the culture of the perpetrators and victims. That is, we assessed not only blame attributions on individual perpetrator and victim level but also on cultural level.

We replicated harsher victim blaming when the perpetrator stemmed from the outgroup. In addition, the ingroup victim was judged more harshly compared to the outgroup victim. Thus, these victim blaming patterns were not due to the intergroup dyad per se.

Mirroring our expectations concerning culture blaming, we observed harsher attributions of blame towards the outgroup culture than the ingroup culture (in other words the ingroup culture was judged less severely than the outgroup culture). That is, on one hand, we observed participants condemning the individual ingroup perpetrator and on the other hand the outgroup culture for the deviant act.

We proceeded to test the protection of the ingroup on individual and on cultural level as part of manuscript 4. We tested whether outgroup saliency before the confrontation with ingroup deviance would come along with harsher black sheep effect due to a distancing strategy from the outgroup. Even though this outgroup saliency manipulation through the quiz questions did not affect the judgments, the results of the proceeding experiments hint to the predicted pattern. When the intergroup context was salient (Exp. 2-3 of manuscript 4), we observed order effects which are in line with the idea that the presence of a threatening outgroup elicits a distancing strategy form the ingroup deviant in form of the black sheep effect. More precisely, when participants worked through the outgroup

newsflash first, they expressed a black sheep effect (i.e., the ingroup perpetrator who came first was judged more harshly). This may be a result of the comparative context (S. A. Haslam & Oakes, 1995) and is discussed further below.

Based on previous literature (Lewis & Sherman, 2010), we further predicted participants in a high ingroup entitativity condition (Campbell, 1958; Crawford et al., 2002; Yzerbyt et al., 2000) expressing a stronger black sheep effect and less ingroup culture blaming than those in the low entitativity condition. We did not observe any main or interaction effect with ingroup entitativity. One possible explanation for this may be that entitativity is difficult to manipulate for social groups. For instance, previous work manipulated entitativity for minimal groups or group of friends (Crawford et al., 2002; Pereira & van Prooijen, 2018) but not for social groups such as Germans. Where entitativity is manipulated for social groups, its impact is not estimated in the context of deviant behavior (cf. Castano et al., 2003).

In sum, in several experiments, we observed a shift of blame, from the stereotyped outgroup perpetrator to the ingroup perpetrator and ingroup victim on individual level. And at cultural level, we predicted, observed and replicated the maintenance of the positivity of the ingroup as a whole.

In short: questions answered as part of the present work

Considering the questions addressed as part of the present work (see Table 1 from the introduction of this thesis), we can summarize the following answers. First, German participants are indeed more interested in stereotypic information categories (e.g., religious affiliation, ethnic background) in face of a perpetrator from a predominantly Muslim country than his ingroup counterpart.

Second, the categorization of the perpetrator does not interact with the crime type which is in line with the stereotype of non-German perpetrators including both property and violent crimes (Bolesta et al., 2022). Further, the benefit of the doubt may on some occasions depend on other factors to occur. Reducing infrahumanization did neither affect positive nor negative emotions in the predicted way. However, ascribing higher positive secondary emotions to the outgroup was associated with less severe blame attributions towards the outgroup perpetrator. This hints to an exonerating effect which suggests a shift of blame again away from the outgroup perpetrator.

Alarmingly, when a perpetrator is described as positive (warm and competent), German participants exonerate his behavior. Throughout the experiments, we observed harsher victim blaming (German victim) when the perpetrator stemmed from the outgroup.

Most important, blame attributions on cultural level hint at the exoneration of the ingroup as a whole. The other side of the coin may be the derogation of the outgroup culture (discussed further below). The final manuscript of this work examined and replicated the protection of the ingroup on individual and on cultural level. One main contribution of this last empirical work lies in the importance of comparative context: participants judge an ingroup perpetrator more harshly when the stereotyped outgroup is presented first. These order effects will be discussed more in depth as part of the next section.

"(N)one among us, but all of them" – on order effects

The robust order effects that we observed as part of two experiments point to an interesting explanation for the black sheep effect and culture blaming. Recall that the order effect, which was replicated as part of manuscript 4, Experiment 3, translates to a black sheep effect when the outgroup newsflash came first. This is in line with the idea that outgroup salience before confrontation with ingroup deviance (manuscript 4, Exp. 1) may elicit the black sheep effect. When working through the outgroup newsflash first, participants have been in an intergroup context (S. A. Haslam et al., 1995; S. A. Haslam & Oakes, 1995) which motivated the harsher judgment of the ingroup perpetrator who was presented second probably due to activated stereotypes about the outgroup.

The intergroup context may be a necessary boundary condition for the black sheep effect to occur in a context where the deviant act is stereotypically associated with a given outgroup.

Limitations of the present work and future directions

With all efforts to conduct exhaustive and fruitful experiments, the present work comes along with some limitations. First, we focused on German samples as ingroup and perpetrators from predominantly Muslim countries as outgroup. For generalization purposes beyond these two groups, one may want to examine blame attributions cross-nationally on perpetrator and cultural level. I would expect the protection of the ingroup on individual (e.g., Marques & Yzerbyt, 1988; Reese et al., 2013) and cultural level to occur cross-nationally whenever an intergroup context with a stigmatized outgroup is at stake and

the ingroup's guilt or hostile intention is not ambiguous (van Prooijen, 2006; Wang et al., 2016).

Second, the infrahumanization manipulation as part of the third contribution of this thesis did not fully work as predicted, however the results hint to a fruitful line of work which benefits from further investigating the role of valence of emotions in infrahumanization theory specially when the manipulated context is incongruent with the assessed valence of the secondary emotions. Recall that following the infrahumanization manipulation participants ascribed fewer negative secondary emotions to the outgroup whereas there was an increase in ascription of the positive secondary emotions. It remains worthwhile further investigating this discrepancy which is not predicted by the theory.

In addition, our experimental manipulations of the fourth contribution of this thesis (outgroup priming and ingroup entitativity), which aimed to increase the protection of the ingroup did not have an impact on the judgments (probably due to ceiling effects). Future work may investigate whether it is possible to experimentally increase the black sheep effect in such deviant contexts with social groups. This would allow to test further underlying mechanisms of ingroup favoritism and outgroup derogation. Furthermore, future work may benefit from including outgroup entitativity as a moderator: as for the outgroup, high entitativity comes along with the so-called interchangeability between outgroup members, indicating that they are all the same (Vasquez et al., 2015) eliciting vicarious retribution. Future work may investigate outgroup entitativity as an underlying factor for the differences in judgments on cultural level.

Third, we manipulated the presentation order of the newsflashes as part of the fourth manuscript, Exp 2-3 but did not examine possible order effects due to the presentation order of the dependent measures. Future work may benefit from this differentiation. It may be fruitful to examine whether the judgments on individual and on cultural level are independent of the presentation order of the assessments. Interestingly, one previous work shows that when participants can first distance themselves from the whole ingroup, then the black sheep effect is not observed anymore (Eidelman & Biernat, 2003). One may ask whether it is necessary to punish the ingroup deviant more harshly on individual level when one previously had the chance of expressing that the ingroup culture is less responsible for the deviant behavior than the outgroup culture. In other words, perceiving the outgroup culture as more responsible than the ingroup culture may suffice on cultural level and the ingroup favoritism may vanish on individual level. However, one may predict both patterns even when the cultural items are presented first and the perpetrator items second. This may

be due to possible different underlying mechanisms (see further below the lack of correlation between perpetrator blaming and culture blaming) which go beyond what was observed as part of the present work. If the culture blaming translates more to outgroup culture derogation rather than ingroup protection per se, then the necessity of preserving the positivity of the ingroup (Marques, 1990; Marques & Yzerbyt, 1988) exists on individual level, as such the black sheep effect shall be expected independent of the presentation order of the cultural items.

Fourth, we investigated negative deviant acts across all experiments. It may be interesting to examine whether the discrepancies would occur also for positive acts achieved or committed by in- and outgroup members and in how far their culture is at stake for the positive deeds.

Strengths of the present work

The present dissertation comes along with several strengths which are mentioned in the following. We conducted several large sampled experiments while trying to reach the general public, that is, we used several platforms for distributing our calls for participation (samples of manuscript 1-3). As such, although our sample focusses on only one ethnicity as social ingroup, that is, Germans, our results are not limited to a psychology student sample as has been criticized for several psychological experiments (cf. Henrich et al., 2010).

It is further important to note, that we tried to reach ecological validity (Kihlstrom, 2021) by providing participants with alleged newspaper articles which are close to media reports in real life. Further, we used social groups (beyond for instance minimal groups) which is also increasing the ecological validity of our set of experiments. We operationalized the outgroup using names, nationalities or referred to a possible religious affiliation: perpetrator from a predominantly Muslim country (manuscript 1), Syrian refugee (manuscript 2), Afghan refugee (manuscript 3, experiments 1-2 in the positive and negative description condition, in the no description condition only the name was manipulated), Afghan perpetrator (manuscript 4). The observed outgroup culture blaming confirms the associations towards a conglomerate of perpetrator from predominantly Muslim countries and refugees (cf. Shooman, 2012; Stürmer et al., 2019; Wigger, 2019). This may allow to generalize the findings beyond the experimental setting.

Where the experimental manipulation (for instance infrahumanization) did not come along with the expected pattern, the correlative patterns indicated interesting results which

were discussed in terms of future directions (e.g., in case of the ascribed secondary positive emotions towards the outgroup, recall the results of manuscript 2, Exp. 2).

Additionally, we used and investigated both experimental and correlative patterns as part of our experiments. The categorization of the perpetrators via name and ethnicity led to robust effects, which were replicated across experiments (e.g., black sheep effect, victim blaming when perpetrator stemmed from the outgroup, culture blaming). We further aimed to achieve maximal parallelism between experimental conditions. This was for instance achieved by keeping parallel all information except for the information revealing the ethnicity of perpetrators and victims.

Further, it is noteworthy, that we aimed to conduct follow-up experiments in a way that allowed to test the replication of at least one previous condition. For instance, even though we did not replicate the findings observed by van Prooijen (2006) as part of manuscript 2, Exp. 1, we conducted a design as part of the follow up Experiment which allowed to examine the effect of guilt certainty again. Another example in this sense targets the examination of the order effects as part of manuscript 4, experiments 2 and 3. This is important as we aimed to test for the robustness of our findings given the replication crisis in psychology (e.g., Shrout & Rodgers, 2018).

This is also the first contribution investigating differential patterns in degree of interest towards information categories in the context of two crime types (theft and sexual violence) while taking into account the group membership of the perpetrator.

We further investigated differences regarding judgments on individual perpetrator and victim as well as cultural level. We observed discrepancies on individual and on cultural level which would have been undetected if we would have investigated only the judgments on one of these two levels. We observed the protection of the ingroup both on individual and cultural level while going beyond allocations of positive or negative stimuli (e.g., Mummendey et al., 2000; Mummendey & Otten, 1998) or a costly punishment game (cf. Schiller et al., 2014). In our experiments we examined guilt attributions in a highly deviant context (e.g., sexual violence) which has been in the spotlight of the media and debates about integration (Landmann et al., 2019; Stürmer et al., 2019; Wigger, 2019; Wigger et al., 2022). As such we looked at for instance what is associated with safety threat or realistic threat situations (Landmann et al., 2019; Stephan & Stephan, 1996, 2000). The present work provides empirical hints to ingroup favoritism on individual level (in form of the black sheep effect) and on cultural level which may be translated to the derogation of the

stereotyped outgroup. I will discuss these two sides of one coin in the following more in depth.

The protection of the ingroup on individual and cultural level – an extension to a mere black sheep effect?

On the one hand we observed the black sheep effect as part of several experiments (manuscript 2, partly in Exp. 1, Exp. 2, manuscript 3, Exp. 1-2, manuscript 4, Exp. 2-3) and on the other hand, we observed more lenient judgments towards the ingroup than outgroup culture in all experiments where culture blaming was assessed (manuscript 3, Exp. 2, manuscript 4, Exp. 1-3). One may interpret the data as expressing more leniency towards one outgroup deviant but harsher derogation of the outgroup as a whole on cultural terms. So far, we do not fully know whether the data supports ingroup favoritism on individual and cultural level or whether we have ingroup favoritism on individual level and outgroup derogation on cultural level. The following paragraphs aim to provide possible answers for these questions.

Beyond differentiating from the outgroup?

One perspective from previous work on ingroup bias and outgroup discrimination hints to the pattern that ingroup bias aims at differentiating from outgroups. This work illustrates that outgroup derogation can take place when there is no alternative way for ingroup positivity than in expense of the outgroup (Mummendey & Schreiber, 1983). This was investigated by the use of different questionnaires for assessing the quality of a creative group work. Ingroup bias was highest in the condition where the assessment was complementary and enabled participants to distribute 100 points between the in- and outgroup. In this condition one group may have received higher ratings than the other group as the points were possible to benefit one group over the other. The authors showed that in- and outgroups are judged equally when assessment takes place for the same dimensions and separately for both groups.

We argue that in our experiments the assessments for both groups were also on equal dimensions and separate (same dependent measures: individual and culture for both in- and outgroup perpetrator), thus in principle enabling equal treatments for participants. However, we observed discrepancies on individual and cultural judgments (individual judgments ingroup > outgroup, culture blaming ingroup < outgroup). Reasons for this discrepancy may lie in differences regarding the methods used. For instance, we looked at

negative deviant acts while participants of Mummendey and Schreiber (1983) evaluated a creative group work where groups of participants had to come up with a painting of a "human city of tomorrow". Further, groups were created by using minimal groups (participants differed in the color of the smocks they wear). In addition, the participants were 66 male subjects from a college of social work. One may speculate whether social work students may be more lenient regarding outgroup treatment thus showing outgroup discrimination only when it was not possible to secure ingroup positivity without damaging the outgroup. More importantly, in our work, we show outgroup discrimination in the context of a negative deviant behavior and examined the blame attributions towards both individual members as well as on cultural level for real social categories. We can conclude that considering the differences between Mummendey and Schreiber's work (1983) and the present work, we indeed observed discrimination against the outgroup on cultural terms while using separate assessments. As such, using separate assessments may in some contexts reveal outgroup discrimination which may go beyond mere differentiation from the outgroup. Alternatively, the reader may ask about the correlations between the assessments (individual and culture blaming), as a negative correlation may indicate that the assessments have not been separate and that the higher the judgments of the individual ingroup deviant, the more lenient the judgments on cultural level. This was exploratively tested for manuscript 4, Experiment 2 and 3, where we can directly compare the correlation patterns as part of the full within-subjects framework as well as part of the between-subjects framework (first presented newsflash). To examine the correlation between perpetrator blaming and culture blaming in the full within-subjects framework, we created two difference scores (ingroup minus outgroup) for perpetrator blaming and culture blaming each. This correlation between these two difference scores was not significant, $r = .062$, $p = .294$, for Experiment 2 and $r = .002$, $p = .963$, for Experiment 3 of manuscript 4. We further looked at the correlation pattern between perpetrator blaming and culture blaming while only looking at the first presented newsflash. Here, again we observed no meaningful correlation between perpetrator blaming and outgroup culture blaming $r = -.022$, $p = .708$, for Experiment 2, first newsflash only, $r = -.089$, $p = .019$ for Experiment 3, first newsflash only. Neglecting the order of presentation, we did not observe any correlation between ingroup perpetrator blaming and outgroup culture blaming $r = -.013$, $p = .830$ for Experiment 2 and $r = -.051$, $p = .175$ for Experiment 3 of manuscript 4.

The absence of a correlation between the two is reminiscent of previous research arguing that ingroup love and outgroup discrimination do not relate to each other and have

differing underlying mechanisms (Brewer, 1979, 1999, 2001, 2017; Brewer & Campbell, 1976; Cashdan, 2001; Hewstone et al., 2002; Struch & Schwartz, 1989; Yzerbyt & Demoulin, 2010). As such the lack of a correlation is not surprising in our framework. A further analysis of additional data provides insights for outgroup prejudice driving the difference in culture blaming. Here, as part of the broader project of manuscript 4, Experiment 1 we assessed negative attitudes towards the outgroup two weeks before participants were invited to the main Experiment. Interestingly, with increasing negative attitudes towards Muslims and Islam participants indicated harsher derogation of the outgroup culture. Thus, we can conclude that prejudice moderated the derogation of the outgroup culture[24].

Considering attributional differences on two different assessments, the reader may also think of work on shifting standards. As part of the following section, I will briefly relate the present work with the shifting standards literature.

Shifting standards

Another explanation for the patterns observed (black sheep effect on individual level and the harsher judgments towards the outgroup culture) may be due to different standards based on stereotypes people endorse towards other groups (Biernat et al., 1991; Linville & Jones, 1980). For instance, an underlying stereotype based on threats associated with refugees and Muslims may read as follows "*Muslims/refugees are dangerous or violent*" (Frissen et al., 2018; Landmann et al., 2019; Velasco González et al., 2008; Wirtz et al., 2016) and lead to differential judgments. One may perceive the black sheep effect as a

[24] *Culture blaming*. The regression analysis with perpetrator culture blaming as criterion indicated the predicted main effect of couple, $b = .36$, $t(426) = 2.39$, $p = .017$, $\eta_p^2 = .01$, such that participants blamed the outgroup culture more strongly ($M = 3.27$, $SD = 1.77$) than the ingroup culture ($M = 2.89$, $SD = 1.56$). We further observed a significant negative attitudes towards the outgroup (NA) effect, $b = .57$, $t(426) = 6.03$, $p < .001$, $\eta_p^2 = .08$, and a significant couple × NA interaction, $b = 1.28$, $t(426) = 6.78$, $p < .001$, $\eta_p^2 = .10$. To decompose this interaction, we ran simple slope analyses as a function of couple. Participants blamed the culture of the outgroup more with increasing NA, $b = 1.21$, $t(426) = 9.13$, $p < .001$, $\eta_p^2 = .164$, while there was no relationship between NA and blame attributed to the ingroup culture, $p = .600$

coping strategy with the violated stereotype (Carnaghi & Yzerbyt, 2007; Pinquart et al., 2021). Within the ingroup category it is highly negative to act as the ingroup deviant in our experiments. As such the black sheep effect may occur on individual level as the ingroup deviant acted contradictory to the general expectancy, in other words the ingroup is evaluated as more harshly as the standards for the ingroup are more demanding (Biernat, 2003; Biernat et al., 1991). In case of the outgroup perpetrator the above mentioned prejudice may be confirmed which may explain the harsher derogation of the outgroup than ingroup culture (Biernat et al., 1991; Biernat & Manis, 1994; Linville & Jones, 1980). As such we could translate shifting standards to the two levels of individual and cultural level of the present work. Even though the concept of shifting standards was shown using different dependent measures which were beyond the scope of the present work, we perceive similarities regarding the differential judgments (Alves, 2022, personal communication). Different cognitive standards about the ingroup and outgroup may be in line with the observed judgments on individual and cultural level. Further above we observed the moderating impact of negative attitudes towards the outgroup, on culture blaming: With increasing negative attitudes towards Muslims and Islam, participants more harshly blamed the outgroup than the ingroup.

Bringing in negative attitudes in the explanation framework, point to a motivational desire behind the observed patterns. Interestingly, a motivational explanation for shifting standards has been provided as well (Miron et al., 2010). The motivation to protect the ingroup can on the one hand drive to stricter standards leading to the protection of the ingroup and on the other hand to less collective blame for the ingroup (Miron et al., 2010). Miron et al. (2010) investigated people's reactions towards the African colonization by America. High identifiers set the standards for the judgment of the negative past of the ingroup higher than low identifiers which resulted in less severe judgments of the harm as well as less perceived ingroup collective guilt (Miron et al., 2010). In other words, high identifiers needed more evidence for accepting their ingroups' wrongdoing which protected from high perceptions of collective guilt. Thus, shifting standards are not only a cognitive mechanism as shifting standards can also be understood as being caused by motivational desires (Miron et al., 2010). However, these results do not fully overlap with the black sheep effect on individual level as here high identifiers perceived the harm as less severe and thus experienced less collective guilt. In our experiments participants judged the ingroup deviant more harshly and exonerated the ingroup culture. But it is important to mention that in our experiments we examined both in- and outgroup wrongdoing while

Miron et al. (2010) investigated ingroups' wrongdoing (for instance, Exp. 1, Americans enslaving Africans). However, both perceiving a harm as less severe or harsher rejecting it parallel to the two reactions illustrated as part of the introduction of this thesis, namely, positive or negative bias as part of the coping with ingroup deviance model. As such both Miron et al. (2010) and our work may be interpreted in line with a motivational account of preserving the positivity of the ingroup, while as part of the present thesis, we investigated both in- and outgroup wrongdoing and in addition examined blame judgments on individual and cultural level, which contributes to the existing literature on intergroup attributions.

Future work may benefit from implementing an experimental manipulation of Miron et al. (2010, Study 3) to examine whether the judgments on individual and on cultural level may be influenced by activating group-affirmation. More precisely, the authors asked their participants to list three positive things that the ingroup did as part of the history, for reducing the motivation to protect the ingroup. The positive affirmation condition decreased the motivation of setting higher ingroup standards for concluding that the ingroup has been racist. Participants of this condition also perceived the harm caused by the ingroup as more severe and expressed higher collective ingroup guilt than those of the non-positive affirmation condition (Miron et al., 2010). It would be fruitful to investigate whether such a group-affirmation task beforehand may lead to more or less equal cultural blame attributions which can in turn be used for interventions for reducing discrimination of the outgroup on cultural level.

It is important to note, that in line with our patterns regarding the lack of a correlation between perpetrator and culture blaming, no correlation has been observed between shifting standards and other forms of prejudice (direct: explicit racial attitudes nor indirect: via Implicit association test) in the context of academic competence of Black and White people (Biernat et al., 2009). However, shifting standards predicted behavioral discrimination, that is, decreased allocation of funds towards an organization in support of the outgroup (Biernat et al., 2009). Both shifting standards and the patterns observed as part of the present dissertation on individual and cultural level may be based on stereotypes. The comparative judgment (ingroup < outgroup) in culture blaming can be understood as a cultural derogation, as a form of discrimination of the whole outgroup. If what we observed on cultural level translates to an outgroup discrimination and taking into account the black sheep effect which preserves the positivity of the ingroup by excluding one ingroup deviant, the present work provides empirical data for what Brewer classified as

one of her typologies, which I will discuss in the next section together with different predictors underlying ingroup favoritism and outgroup derogation.

Ingroup favoritism and outgroup discrimination: Brewer's Type III typology

Indeed, recent research points to different predictors for ingroup favoritism and outgroup derogation for in- and outgroup warmth ratings (Hamley et al., 2020). Using latent profile analysis, researchers observed social dominance orientation, preference for inequality and social hierarchies among groups (Pratto et al., 1994), predicting ingroup favoritism and outgroup derogation for Europeans, that is, the advantaged group (Hamley et al., 2020). For the disadvantaged group (i.e., Māori, minority in New Zealand) ingroup favoritism and outgroup derogation was predicted by ethnic identity centrality (Hamley et al., 2020).

What Hamley et al. (2020) observe based on their analyses, reflects Type III of Brewer's typology of ingroup bias decomposed by Brewer (2017): Type I reflects ingroup favoritism which is expressed without any disadvantageous treatment of the outgroup (Brewer, 2017). Type II translates to discrimination of the outgroup while the ingroup's treatment is unbiased. Finally, Type III translates to what we may have observed as part of our empirical work: "discrimination involves differential treatment in favor of the ingroup *and* against the outgroup"(Brewer, 2017 p. 92). Type III includes threats which are elicited by specific outgroups and thus this ingroup bias aims to protect the ingroup (Brewer, 2017). Brewer (2017) points in her review chapter that most research centers around the first Type of ingroup bias (i.e., here the ingroup is favored while there is no hostility towards the outgroup). The present work points to Type III typology of ingroup favouritism which includes the discrimination of the outgroup in form of the outgroup culture derogation. The further above-mentioned moderation effect between perpetrator's group membership and negative attitudes towards the outgroup on culture blaming (outgroup > ingroup) indicated the underlying prejudice for the outgroup culture derogation. As part of the next section, I will briefly turn to the question whether an unaffiliated group would be necessary for differentiating between ingroup favoritism and outgroup derogation.

The (un)necessity of an unaffiliated group?

The interested reader may ask whether the present design of experiments lack a control condition where an unaffiliated perpetrator is compared to both ingroup and outgroup perpetrator. In theory, this may allow to differentiate between the underlying

processes, for instance whether it is indeed more ingroup protection or outgroup derogation on individual and on cultural terms. Creating such an unaffiliated group is indeed challenging as we cannot be sure how the imagination of participants will perceive an alleged unaffiliated perpetrator (as unaffiliated, ingroup or outgroup?). In real world, there are indeed conflicting groups which face each other, as such one may argue that the present work is close to the real-world scenario. However, previous work hint to the importance of differentiating between the two processes as most of work in intergroup context do not investigate a third unaffiliated group (Braun & Gollwitzer, 2012). As such future work may benefit from investigating this further by implementing a possible unaffiliated group (Khosrowtaj et al., 2023). All in all, it is important to note that the black sheep effect is a comparative judgment where the ingroup is evaluated more extremely compared to an outgroup. For culture blaming, we argue that it is as well a comparative judgment which disadvantages the outgroup over the ingroup. As further analysis of additional data revealed that negative attitudes towards the outgroup impact the outgroup culture blaming, I believe that the difference in culture blaming is a cause of outgroup derogation rather than mere ingroup protection. Future work can elaborate on these underlying processes more in depth.

Conclusions

The introduction of this work started with a quote hinting to the loss of individuality of outgroups such as refugees. The present dissertation examined several characteristics affecting the coping with ingroup deviance, while extending the ingroup deviance model with investigating reactions towards in- and outgroup deviants on individual and on cultural level. Alarmingly, stereotypic information categories such as religious background and ethnic background of the outgroup perpetrator marked interesting information categories which participants wished to know further about.

Regarding the characteristics affecting blame attributions, we can conclude that it matters whether a victim is harmed by an outgroup or an ingroup perpetrator. In case of an outgroup perpetrator the victim is judged more harshly. Compared to an outgroup victim, the ingroup victim is perceived as more blameworthy. Positive descriptions regarding warmth and competence lead to the exoneration of a perpetrator. Humanization (correlative pattern only for positive secondary emotions) is associated with less severe judgments of an outgroup perpetrator.

In line with stereotypes towards refugees, Muslims and foreigners (Frissen et al., 2018; Ogan et al., 2014; Soral et al., 2020; Stürmer et al., 2019; Unkelbach et al., 2008;

Velasco González et al., 2008; Wigger, 2019; Wigger et al., 2022) as well as the black sheep literature (Marques et al., 1988; Marques & Yzerbyt, 1988), we tested the protection of the ingroup both on individual and cultural level. Participants judged one individual perpetrator more harshly (black sheep effect) while on cultural level the outgroup culture was perceived as more responsible for the deviant act than the ingroup culture (ingroup < outgroup). This may be the empirical hint for the loss of individuality which was illustrated by the quote described as part of the introduction of the present thesis (Gouma, 2019). One outgroup deviant may be not punished as harshly as his ingroup counterpart because he is perceived to represent a whole group. The culture blaming pattern is reminiscent of a subtle form of prejudice:

"*Note that exaggerated cultural differences are ostensibly non-prejudicial – the covert key to prejudice; for it hardly appears prejudiced to report on actual and obvious intergroup differences"* (Pettigrew & Meertens, 1995 p. 60).

This possible culture derogation may have remained invisible without having examined the attributions beyond the individual perpetrators. As such for future work, we also suggest differentiating on attributional levels (referring to one acting target and their culture). Future work may for instance follow this fruitful path by (qualitatively) examining participants' perceptions of discrepancies in responsibility attributions which are presumably taken as *cultural differences*. The outgroup may thus be perceived as an enduring homogenous threat and the positivity of the ingroup as a whole is preserved (Marques, 1990; Marques et al., 1988; Marques & Yzerbyt, 1988; Yzerbyt et al., 2000).

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), Article 2. https://doi.org/10.1037/rev0000262

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), Article 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

Ahmed, S., & Matthes, J. (2017). Media representation of Muslims and Islam from 2000 to 2015: A meta-analysis. *International Communication Gazette*, *79*(3), Article 3. https://doi.org/10.1177/1748048516656305

Bastian, B., Denson, T. F., & Haslam, N. (2013). The Roles of Dehumanization and Moral Outrage in Retributive Justice. *PLOS ONE*, *8*(4), e61842. https://doi.org/10.1371/journal.pone.0061842

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. (2003). Toward a Broader View of Social Stereotyping. *American Psychologist*, *58*, 1019–1027. https://doi.org/10.1037/0003-066X.58.12.1019

Biernat, M., Collins, E. C., Katzarska-Miller, I., & Thompson, E. R. (2009). Race-Based Shifting Standards and Racial Discrimination. *Personality and Social Psychology Bulletin*, *35*(1), 16–28. https://doi.org/10.1177/0146167208325195

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

Biernat, M., Vescio, T. K., & Billings, L. S. (1999). Black sheep and expectancy violation: Integrating two models of social judgment. *European Journal of Social Psychology*, *29*, 523–542. https://doi.org/10.1002/(SICI)1099-0992(199906)29:4<523::AID-EJSP944>3.0.CO;2-J

Bleiker, R., Campbell, D., Hutchison, E., & Nicholson, X. (2013). The visual dehumanisation of refugees. *Australian Journal of Political Science*, *48*(4), 398–416. https://doi.org/10.1080/10361146.2013.840769

Bolesta, D., Oehme, & Führer, J. L. (2022). *Stereotype über Straftäter:innen* (Panel Zur Wahrnehmung von Kriminalität Und Straftäter:Innen (PaWaKS): Ergebnisse Der Ersten Erhebungswelle. Zentrum Für Kriminologische Forschung Sachsen e.V.). https://www.zkfs.de/pawaks/

Braun, J., & Gollwitzer, M. (2012). Leniency for out-group offenders. *European Journal of Social Psychology*, *42*(7), Article 7. https://doi.org/10.1002/ejsp.1908

Brems, C., & Wagner, P. (1994). Blame of Victim and Perpetrator in Rape Versus Theft. *Journal of Social Psychology*, *134*(3), Article 3. https://doi.org/10.1080/00224545.1994.9711741

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), Article 2. https://doi.org/10.1037/0033-2909.86.2.307

Brewer, M. B. (1999). The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues*, *55*(3), Article 3. https://doi.org/10.1111/0022-4537.00126

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.

Brewer, M. B. (2017). Intergroup discrimination: Ingroup love or outgroup hate? In *The Cambridge handbook of the psychology of prejudice* (pp. 90–110). Cambridge University Press. https://doi.org/10.1017/9781316161579.005

Brewer, M. B., & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence* (p. 218). Sage.

Bruneau, E., Kteily, N., & Laustsen, L. (2018). The unique effects of blatant dehumanization on attitudes and behavior towards Muslim refugees during the European 'refugee crisis' across four countries. *European Journal of Social*

Psychology, *48*(5), Article 5. https://doi.org/10.1002/ejsp.2357

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

Cashdan, E. (2001). Ethnocentrism and Xenophobia: A Cross-Cultural Study. *Current Anthropology*, *42*(5), 760–765. https://doi.org/10.1086/323821

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infrahumanization in response to collective responsibility for intergroup killing. *Journal of Personality and Social Psychology*, *90*(5), Article 5. https://doi.org/10.1037/0022-3514.90.5.804

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

Castano, E., & Yzerbyt, V. Y. (1998). The highs and lows of group homogeneity. *Behavioural Processes*, *42*, 219–238. https://doi.org/10.1016/S0376-6357(97)00078-8

Chadwick. (2015, October 19). The top 10 wackiest anti-refugee remarks. *POLITICO*. https://www.politico.eu/article/toxic-news-refugees-migrants-eu/

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

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), Article 1. https://doi.org/10.1348/014466608X314935

Cuddy, A. J. C., Rock, M. S., & Norton, M. I. (2007). Aid in the Aftermath of Hurricane Katrina: Inferences of Secondary Emotions and Intergroup Helping. *Group Processes & Intergroup Relations*, *10*(1), Article 1. https://doi.org/10.1177/1368430207071344

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

Davies, T., Yogeeswaran, K., Verkuyten, M., & Loughnan, S. (2018). From humanitarian aid to humanization: When outgroup, but not ingroup, helping increases humanization. *PLOS ONE*, *13*(11), Article 11. https://doi.org/10.1371/journal.pone.0207343

Demoulin, S., Cortes, B., Viki, T., Rodríguez, A., Rodriguez, R., Paladino, M., & Leyens, J.-P. (2009). The Role of In-group Identification in Infra-humanization. *International Journal of Psychology : Journal International de Psychologie*, *44*, 4–11. https://doi.org/10.1080/00207590802057654

Donovan, R. A. (2007). To Blame or Not To Blame: Influences of Target Race and Observer Sex on Rape Blame Attribution. *Journal of Interpersonal Violence*, *22*(6), Article 6. https://doi.org/10.1177/0886260507300754

Doosje, B., Zebel, S., Scheermeijer, M., & Mathyi, P. (2007). Attributions of Responsibility for Terrorist Attacks: The Role of Group Membership and Identification. *International Journal of Conflict and Violence (IJCV)*, *1*(2), Article 2. https://doi.org/10.4119/ijcv-2749

Duxbury, S. W., Frizzell, L. C., & Lindsay, S. L. (2018). Mental Illness, the Media, and the Moral Politics of Mass Violence: The Role of Race in Mass Shootings Coverage. *Journal of Research in Crime and Delinquency*, *55*(6), Article 6. https://doi.org/10.1177/0022427818787225

Eidelman, S., & Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology*, *39*(6), Article 6. https://doi.org/10.1016/S0022-1031(03)00042-8

Ellemers, N., & Haslam, S. A. (2012). Social identity theory. In *Handbook of theories of social psychology, Vol. 2* (pp. 379–398). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n45

Esses, V. M., Medianu, S., & Lawson, A. S. (2013). Uncertainty, Threat, and the Role of the Media in Promoting the Dehumanization of Immigrants and Refugees. *Journal of Social Issues*, *69*(3), 518–536. https://doi.org/10.1111/josi.12027

Fabregat, E., Medina-Bravo, P., & Meyers, M. (2019). NEIGHBORHOOD PRESS AND IMMIGRATION A frame-building analysis of the depiction of immigrants.

Journalism Practice, *13*(5), Article 5. https://doi.org/10.1080/17512786.2018.1521705

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), Article 2. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), Article 6. https://doi.org/10.1037/0022-3514.82.6.878

Frissen, T., Coppens, T., Tobback, L., & d'Haenens, L. (2018). From worrying about terror to being convinced of the 'Muslim danger. A cross-sectional investigation of the association between exposure to terrorism news coverage, terrorism catastrophizing, and threat related attitudes towards Muslims. *Tijdschrift Voor Communicatiewetenschap*, *46*(3), Article 3. https://www.webofscience.com/wos/woscc/full-record/WOS:000448495500005

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

Gleason, J. M., & Harris, V. A. (1976). Group Discussion and Defendant's Socio-Economic Status as Determinants of Judgments by Simulated Jurors1. *Journal of Applied Social Psychology*, *6*(2), Article 2. https://doi.org/10.1111/j.1559-1816.1976.tb01323.x

Gouma, V. (2019, January 28). Lesermeinung: 'Ich bin die Flüchtlinge!' *Der Tagesspiegel Online*. https://www.tagesspiegel.de/gesellschaft/ich-bin-die-fluchtlinge-5314561.html

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

Hamley, L., Houkamau, C. A., Osborne, D., Barlow, F. K., & Sibley, C. G. (2020). Ingroup Love or Outgroup Hate (or Both)? Mapping Distinct Bias Profiles in the Population. *Personality and Social Psychology Bulletin*, *46*(2), Article 2. https://doi.org/10.1177/0146167219845919

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*, 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Haslam, N., & Stratemeyer, M. (2016). Recent research on dehumanization. *Current Opinion in Psychology*, *11*, 25–29. https://doi.org/10.1016/j.copsyc.2016.03.009

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), Article 4. 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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hestermann, T. (2018). Refugees and Migrants in the Media: The Black Hole. In *Refugees and Migrants in Law and Policy: Challenges and Opportunities for Global Civic Education* (pp. 125–136). https://doi.org/10.1007/978-3-319-72159-0_5

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

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Intimate partner and sexual violence against women: Evidence brief. (2019). World Health Organization. https://www.who.int/publications-detail-redirect/WHO-RHR-19.16

Kardos, P., Leidner, B., Castano, E., & Lickel, B. (2019). The benefits of collective

responsibility: How ingroup reputation concern motivates prosociality in intergroup contexts. *European Journal of Social Psychology*, *49*(1), Article 1. https://doi.org/10.1002/ejsp.2506

Karelaia, N., & Keck, S. (2013). When deviant leaders are punished more than non-leaders: The role of deviance severity. *Journal of Experimental Social Psychology*, *49*, 783–796. https://doi.org/10.1016/j.jesp.2013.04.003

Kauff, M. (2022). Muslim = Terrorist? Attribution of violent crimes to terrorism or mental health problems depend on perpetrators' religious background. *The Journal of Social Psychology*, *0*(0), Article 0. 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

Kelman. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues*, *29*(4), Article 4.

Khosrowtaj, Z., Süssenbach, P., Yzerbyt, V., & Teige-Mocigemba, S. (2023). Does the name reveal the blame? Examining the effect of social categorization on blame attributions—A Registered Report. *Manuscript Submitted*.

Kihlstrom, J. F. (2021). Ecological Validity and 'Ecological Validity'. *Perspectives on Psychological Science*, *16*(2), 466–471. https://doi.org/10.1177/1745691620966791

Kotzur, P. F., Friehs, M.-T., Asbrock, F., & Zalk, M. H. W. van. (2019). Stereotype content of refugee subgroups in Germany. *European Journal of Social Psychology*, *0*(0), Article 0. https://doi.org/10.1002/ejsp.2585

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

Kteily, N. S., & Landry, A. P. (2022). Dehumanization: Trends, insights, and challenges. *Trends in Cognitive Sciences*, *26*, 222–240. https://doi.org/10.1016/j.tics.2021.12.003

Landmann, H., Gaschler, R., & Rohmann, A. (2019). What is threatening about refugees? Identifying different types of threat and their association with emotional responses and attitudes towards refugee migration. *European Journal of Social Psychology*, *49*(7), Article 7. https://doi.org/10.1002/ejsp.2593

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

Leyens, J.-P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The Emotional Side of Prejudice: The Attribution of Secondary Emotions to Ingroups and Outgroups. *Personality and Social Psychology Review*, *4*(2), Article 2. https://doi.org/10.1207/S15327957PSPR0402_06

Leyens, J.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.-P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, *31*(4), 395–411. https://doi.org/10.1002/ejsp.50

Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., & Schmader, T. (2006). Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review*, *10*(4), Article 4. https://doi.org/10.1207/s15327957pspr1004_6

Lickel, B., Schmader, T., & Hamilton, D. L. (2003). A Case of Collective Responsibility: Who Else Was to Blame for the Columbine High School Shootings? *Personality and Social Psychology Bulletin*, *29*(2), 194–204. https://doi.org/10.1177/0146167202239045

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

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., & Paez, D. (1994). The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. *European Review of Social Psychology*, *5*(1), Article 1. https://doi.org/10.1080/14792779543000011

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

Masson, T., & Barth, M. (2019). Solving the paradox—(Further) evidence for a quadratic relationship between in-group centrality and group-based guilt. *The British Journal of Social Psychology*, *58*(4), Article 4. https://doi.org/10.1111/bjso.12310

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

Moss, A. J., Blodorn, A., Van Camp, A. R., & O'Brien, L. T. (2019). Gender equality, value violations, and prejudice toward Muslims. *Group Processes & Intergroup Relations*, *22*(2), Article 2. https://doi.org/10.1177/1368430217716751

Mummendey, A., & Otten, S. (1998). Positive-negative asymmetry in social discrimination. In *European Review of Social Psychology* (Vol. 9, pp. 107–143).

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-Negative Asymmetry in Social Discrimination: Valence of Evaluation and Salience of Categorization. *Personality and Social Psychology Bulletin*, *26*(10), 1258–1270. https://doi.org/10.1177/0146167200262007

Mummendey, A., & Schreiber, H.-J. (1983). Better or just different? Positive social identity by discrimination against, or by differentiation from outgroups. *European Journal of Social Psychology*, *13*(4), Article 4. https://doi.org/10.1002/ejsp.2420130406

Noack, R. (2015, September 3). Muslims threaten Europe's Christian identity, Hungary's leader says. *Washington Post*. https://www.washingtonpost.com/news/worldviews/wp/2015/09/03/muslims-threaten-europes-christian-identity-hungarys-leader-says/

Noor, M., Kteily, N., Siem, B., & Mazziotta, A. (2019). "Terrorist" or "Mentally Ill": Motivated Biases Rooted in Partisanship Shape Attributions About Violent Actors. *Social Psychological and Personality Science*, *10*(4), Article 4. https://doi.org/10.1177/1948550618764808

Ogan, C., Willnat, L., Pennington, R., & Bashir, M. (2014). The rise of anti-Muslim prejudice: Media and Islamophobia in Europe and the United States. *International*

Communication Gazette, *76*(1), Article 1. https://doi.org/10.1177/1748048513504048

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

Panitz, C., Endres, D., Buchholz, M., Khosrowtaj, Z., Sperl, M. F. J., Mueller, E. M., Schubö, A., Schütz, A. C., Teige-Mocigemba, S., & Pinquart, M. (2021). A Revised Framework for the Investigation of Expectation Update Versus Maintenance in the Context of Expectation Violations: The ViolEx 2.0 Model. *Frontiers in Psychology*, *12*, 5237. https://doi.org/10.3389/fpsyg.2021.726432

Pereira, A., & van Prooijen, J.-W. (2018). Why we sometimes punish the innocent: 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

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*(1), Article 1. https://doi.org/10.1002/ejsp.2420250106

Pinquart, M., Rothers, A., Gollwitzer, M., Khosrowtaj, Z., Pietzsch, M., & Panitz, C. (2021). Predictors of Coping With Expectation Violation: An Integrative Review. *Review of General Psychology*, 10892680211024124. https://doi.org/10.1177/10892680211024123

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*, 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Reese, G. (2013, September 8). *Das schwarze Schaf unter uns – Warum Menschen negativ auf Nestbeschmutzer reagieren*. In-Mind. https://de.in-mind.org/article/das-schwarze-schaf-unter-uns-warum-menschen-negativ-auf-nestbeschmutzer-reagieren

Reese, G., Steffens, M. C., & Jonas, K. J. (2013). When Black Sheep Make Us Think: Information Processing and Devaluation of in- and Outgroup Norm Deviants. *Social Cognition*, *31*(4), Article 4.

Reich, C. M., Pegel, G. A., & Johnson, A. B. (2022). Are Survivors of Sexual Assault Blamed More Than Victims of Other Crimes? *Journal of Interpersonal Violence*, 08862605211037423. https://doi.org/10.1177/08862605211037423

RKI - GBE Berichte—Gesundheitliche Lage der Frauen in Deutschland. (2022, June 30). https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsB/Gesundheitliche_Lage_der_Frauen_2020.html

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

Saleem, M., Prot, S., Anderson, C. A., & Lemieux, A. F. (2017). Exposure to Muslims in Media and Support for Public Policies Harming Muslims. *Communication Research*, *44*(6), Article 6. https://doi.org/10.1177/0093650215619214

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior*, *35*(3), Article 3. https://doi.org/10.1016/j.evolhumbehav.2013.12.006

Schwartz, S. H., & Struch, N. (1989). Values, Stereotypes, and Intergroup Antagonism. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and Prejudice: Changing Conceptions* (pp. 151–167). Springer. https://doi.org/10.1007/978-1-4612-3582-8_7

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

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Siem, B., Mazziotta, A., Barbarino, M.-L., & Rohmann, A. (2017). „Zusammengepfercht wie die Tiere" – Der Einfluss medialer Berichterstattung auf die Entmenschlichung von Geflüchteten. *The Inquisitive Mind*, *2017*(2), Article 2. https://de.in-mind.org/article/zusammengepfercht-wie-die-tiere-der-einfluss-medialer-berichterstattung-auf-die

Soral, W., Liu, J., & Bilewicz, M. (2020). Media of Contempt: Social Media Consumption

Predicts Normative Acceptance of Anti-Muslim Hate Speech and Islamoprejudice. *International Journal of Conflict and Violence (IJCV)*, *14*, 1–13. https://doi.org/10.4119/ijcv-3774

Stasio, V. D., 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*, *0*(0), Article 0. https://doi.org/10.1080/1369183X.2019.1622826

Stephan, W. G., & Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, *41*(3), Article 3. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

Stephan, W. G., & Stephan, C. W. (1996). Predicting prejudice. *International Journal of Intercultural Relations*, *20*(3–4), Article 3–4. https://doi.org/10.1016/0147-1767(96)00026-0

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In *Reducing prejudice and discrimination* (pp. 23–45). Lawrence Erlbaum Associates Publishers.

Struch, N., & Schwartz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from in-group bias. *Journal of Personality and Social Psychology*, *56*, 364–373. https://doi.org/10.1037/0022-3514.56.3.364

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Muslim 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

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Trope, Y., & Thompson, E. P. (1997). Looking for truth in all the wrong places? Asymmetric search of individuating information about stereotyped group members. *Journal of Personality and Social Psychology*, *73*(2), Article 2. https://doi.org/10.1037/0022-3514.73.2.229

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 der Bruggen, M., & Grubb, A. (2014). A review of the literature relating to rape victim blaming: An analysis of the impact of observer and victim characteristics on attribution of blame in rape cases. *Aggression and Violent Behavior*, *19*(5), Article 5. https://doi.org/10.1016/j.avb.2014.07.008

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

Velasco González, K., Verkuyten, M., Weesie, J., & Poppe, E. (2008). Prejudice towards Muslims in The Netherlands: Testing integrated threat theory. *British Journal of Social Psychology*, *47*(4), Article 4. https://doi.org/10.1348/014466608X284443

Viki, G. T., Fullerton, I., Raggett, H., Tait, F., & Wiltshire, S. (2012). The Role of Dehumanization in Attitudes Toward the Social Exclusion and Rehabilitation of Sex Offenders. *Journal of Applied Social Psychology*, *42*(10), 2349–2367. https://doi.org/10.1111/j.1559-1816.2012.00944.x

Viki, G. T., Osgood, D., & Phillips, S. (2013). Dehumanization and self-reported proclivity to torture prisoners of war. *Journal of Experimental Social Psychology*, *49*(3), Article 3. https://doi.org/10.1016/j.jesp.2012.11.006

Violence against women. (2021). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/violence-against-women

Wang, L., Zheng, J., Meng, L., Lu, Q., & Ma, Q. (2016). Ingroup favoritism or the black sheep effect: Perceived intentions modulate subjective responses to aggressive interactions. *Neuroscience Research*, *108*, 46–54. https://doi.org/10.1016/j.neures.2016.01.011

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual violence: '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

Wirtz, C., van der Pligt, J., & Doosje, B. (2016). Negative attitudes toward Muslims in The

Netherlands: The role of symbolic threat, stereotypes, and moral emotions. *Peace and Conflict: Journal of Peace Psychology*, *22*(1), Article 1. https://doi.org/10.1037/pac0000126

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

Yzerbyt, V., & Demoulin, S. (2010). Intergroup relations. In *Handbook of social psychology, Vol. 2, 5th ed* (pp. 1024–1083). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002028

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory Stereotypes in the French-speaking World. *Group Processes & Intergroup Relations*, *8*(3), Article 3. https://doi.org/10.1177/1368430205053944

Zick, A. (2017). Das Vorurteil über Muslime. In P. Antes & R. Ceylan (Eds.), *Muslime in Deutschland: Historische Bestandsaufnahme, aktuelle Entwicklungen und zukünftige Forschungsfragen* (pp. 39–57). Springer Fachmedien. https://doi.org/10.1007/978-3-658-15115-7_3

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

German Curriculum vitae

<Der Lebenslauf ist nicht Bestandteil der Online-Veröffentlichung.>

English Curriculum vitae

<The curriculum vitae is not part of the online publication.>

Eidesstaatliche Erklärung

Hiermit versichere ich, Zahra Khosrowtaj, dass ich meine Dissertation

On the evaluation of intergroup deviance on individual and group level:
"When it is (n)one of us, but all of them."

selbständig ohne unerlaubte Hilfe angefertigt habe, mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient und alle vollständigen oder sinngemäß übernommenen Zitate als solche gekennzeichnet habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 14.04.2023

Zahra Khosrowtaj