Intergroup Emotion Judgments and Empathy: The Influence of Group Category Membership and Group Relations



Jelka Stojanov

St John's College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2023

Supervisor: Brian Parkinson, PhD

Co-supervisor: Geoffrey Bird, PhD

ACKNOWLEDGEMENTS

The past three years have been an adventure and I am finding it difficult to believe that the time has come for me to write this section. Buckle up everyone, very emotional thank yous are ahead of you!

First and foremost, to my primary supervisor, Professor Brian Parkinson, thank you for all your guidance and encouragement over the years. I have been extremely lucky to have a supervisor who engaged in in-depth discussions in order to bring out the best in students' ideas, who had incredible patience for the fact that most students, myself included, are just trying to figure themselves and their professional and personal paths out, and who responded to all my questions and last-minute reference letter requests so promptly and with such care. I would also like to thank my co-supervisor, Professor Geoff Bird, for supporting me across research, career, and administrative matters, for providing me with opportunities and resources which helped me get where I am now and for always being eager to polish ideas to perfection. Working with both of you has undeniably shaped me into a more perceptive researcher.

To Dr Leor Hackel, thank you for making me feel so welcome in LA, you have truly created a wonderful and friendly environment to learn and excel in and the three months I spent in your lab will remain one of the most unforgettable parts of my DPhil. I greatly appreciate every piece of research and career (and travel!) advice you have offered me. I have truly learnt so much from you and I look forward to working together in the future.

To Dr Danielle Shore, thank you for all the warm guidance during the different stages of my degree and for reassuring me that "everything would be alright in the end". Not only have you been a fantastic mentor to me, but you have also helped me develop an idea of what kind of a mentor I would like to be to others.

To Dr Iris Žeželj, Professor Vanja Ković and Dr Danka Purić, thank you for expanding the horizon of my professional dreams and for offering patient guidance during my "formative" undergraduate years. I have been very lucky to have you as my role models. To Professor Kate Nation, thank you for being such a supportive and caring college advisor. My experience at St John's College was greatly enriched by knowing you. To Dr Shawn Rhoads, thank you for the help with computational modelling, you truly strengthened my belief that collaborative research leads to better science.

I would also like to thank those, without whose generous support, my time in Oxford and this thesis would not have been possible. I am immensely grateful to the Rotary Foundation and the Dulverton Trust for supporting my doctoral studies. Karen, Richard, Keith, and many others, thank you for making me feel welcome and for enriching my time in the UK. Your commitment to the greater good and genuine care for others will never cease to inspire me. Karen, English language lacks the words that can fully capture my gratitude to you, you have been by my side since the first day of my DPhil journey, and I am deeply indebted to you for your help. I am also grateful to the Experimental Psychology Society, St John's College, and Association for Psychological Science for providing the funding to put my research ideas into action and disseminate them to larger audiences. To Eri and her family, thank you for supporting me and contributing to making this Oxford adventure possible.

To my parents, Mihail and Vesna Stojanov (aka Miša and Veca, the ultimate power couple!), your infinite support and love was felt 24/7 across the ~2900 kilometres that separated us. Thank you for everything, without you I would not be a person I am today. I am impatiently looking forward to hosting you in the UK in a few months, and to showing you every beautiful hidden corner of Oxford. I da zajedno popijemo jednu domaću kafu!

To Hélio, you are my soulmate, my best friend, and my partner in crime. Thank you for bringing out the best in me and for believing in me, even when I had little belief myself. I cherish every memory we made along the way and look forward to many more down the line. Let's never stop dancing to loud synth-pop (and occasional turbo-folk, sorry, not sorry) and discussing the meaning of life.

To Dijana (aka Dić), we have been best friends since we were 7 and we will be best friends until we are 107 (yes, I stole the phrase from Politikin zabavnik and yes, in case we live longer than 107, we will be best friends then too!), thank you for everything – for every advice, (e-)hug and cat meme during these three years. Looking forward to drinking wine on the riverbank back home. To Sara and Višnja, my favourite 👽 and 🐉, I feel so thankful to have you in my life, thank you for all the advice, for the pump-up songs you continuously provided me with during the three years and for your contagious optimism that rubbed off on me. To Jelisaveta and Saška, thank you for rekindling my fun side and reminding me that life is more than studying. To Stefan and Denis, thank you for having the patience to deal with my little JavaScript bugs that I was too ashamed to post on Stack Overflow. To Mirta and Josip, thank you for being wonderful collaborators and for making participant recruitment in Croatia and Serbia a bit less challenging. To Andreea, Mira, Mirela, Shiwen, and Sylvia, you have made my time in Oxford so enjoyable, and our college shenanigans have a special place in my memory. To my DPhil mates – Gloria, Jonas, Sam, Peter, Amanda, and Yiran – it was wonderful to share a lab with you. I have learnt so much from you and thank you for being a part of my DPhil journey. To members of the SLAC lab – Begüm, Chelsey, Jean, and Yi – thank you for making me feel at home at USC and for providing me with invaluable feedback during lab meetings. To my GB fam, Paloma and Léo, thank you for being wonderful friends, and I can't wait for more travel adventures in the future.

Finally, I would like to thank all participants who devoted their time to my research. I hope that the insights provided in this thesis would make you proud.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **I**

TABLE OF CONTENTS ..... **V**

LIST OF FIGURES ..... **VIIII**

LIST OF TABLES ..... **IX**

ABSTRACT ..... **X**

CHAPTER 1: INTRODUCTION **1**

GENERAL OVERVIEW OF THE CHAPTER 1
GROUPS AS GROUP CATEGORIES 3
Brief overview of the Social Identity Perspective 4
Summary 9
COMPLEMENTING THE STUDY OF GROUP CATEGORIES WITH THE STUDY OF GROUP FEATURES 10
THE INFLUENCE OF GROUPS ON JUDGMENTS OF FACIAL EXPRESSIONS OF EMOTION 15
Categorical emotion recognition 15
The influence of stereotypes and evaluations on categorical emotion recognition 22
Dimensional judgments of facial expressions of emotion 25
Summary 27
THE INFLUENCE OF GROUPS ON EMPATHY AND THE ROLE OF INTERGROUP EMPATHIC NORMS 27
Defining empathy 28
Intergroup empathy (failures) 34
Interventions to encourage intergroup empathy 40
Summary 51
LIMITATIONS OF STUDYING GROUPS AS GROUP CATEGORIES: HOW CAN FEATURE-ORIENTED THINKING ABOUT GROUPS HELP? 52
Examples of feature-oriented thinking in social psychological research 56
Summary 58
OVERVIEW OF THE CURRENT THESIS 59

CHAPTER 2: THE INFLUENCE OF GROUP CATEGORY MEMBERSHIP AND COALITIONAL RELATIONS ON JUDGMENTS OF FACIAL EXPRESSIONS OF EMOTION **62**

ABSTRACT 62
INTRODUCTION 64
Hypotheses 70
EXPERIMENT 1 71
Method 71
Results 89
Discussion 96
EXPERIMENT 2 98
Method 99
Results 100
Discussion 106
EXPERIMENT 3 106

Eye-tracking and visual attention 107
Peripheral physiology and arousal 115
Method 124
Results 135
Discussion 141
GENERAL DISCUSSION 145

CHAPTER 3: PERCEPTIONS OF EMPATHIC NORMS IN INTERGROUP CONTEXTS .. 153

ABSTRACT 153
INTRODUCTION 155
Hypotheses 158
EXPERIMENT 4 162
Method 162
Results 167
Discussion 178
EXPERIMENT 5 180
Method 180
Results 183
Discussion 197
EXPERIMENT 6A 199
Method 200
Results 202
Discussion 203
GENERAL DISCUSSION 204
Open practices 212

CHAPTER 4: LEARNING TO EMPATHISE: HOW DO INTERGROUP EMPATHIC NORMS INFLUENCE OUTGROUP-ORIENTED EMPATHY? 213

ABSTRACT 213
INTRODUCTION 215
EXPERIMENT 6B 216
Method 219
Results 224
Discussion 226
EXPERIMENT 7 231
Hypotheses 239
Method 244
Results 254
Discussion 273
GENERAL DISCUSSION 288
Open practices 292

CHAPTER 5: GENERAL DISCUSSION 293

GENERAL OVERVIEW OF THE CHAPTER 293
THE INFLUENCE OF GROUP CATEGORY MEMBERSHIP AND COALITIONAL RELATIONS ON JUDGMENTS OF FACIAL EXPRESSIONS OF EMOTION 295
Findings 295
Implications and future directions 298

PERCEPTIONS OF EMPATHIC NORMS IN INTERGROUP CONTEXTS 301
Findings 302
Implications and future directions 305
LEARNING TO EMPATHISE: HOW DO INTERGROUP EMPATHIC NORMS INFLUENCE OUTGROUP-ORIENTED EMPATHY? 308
Findings 309
Implications and future directions 313
BROADER THEORETICAL CONSIDERATIONS AND FUTURE DIRECTIONS 319
Intergroup empathy versus intergroup compassion 319
Relational social categorisation 322
Feature-oriented thinking about groups 324
CONCLUSION 326
REFERENCES 327
APPENDICES 409
CHAPTER 2: SUPPLEMENTARY MATERIALS 409
CHAPTER 3: SUPPLEMENTARY MATERIALS 412
CHAPTER 4: SUPPLEMENTARY MATERIALS 431

LIST OF FIGURES

Figure 1.1. Visualisation of the two discussed views of empathy 32
Figure 2.1. Example stimuli used in Experiment 1 74
Figure 2.2. The shooting game used to induce expectations of cooperation and competition 77
Figure 2.3. The emotion task (online procedure) 81
Figure 2.4. The emotion task (lab-based procedure) 84
Figure 2.5. Intra- and intergroup coalitional preferences (Experiment 1) 88
Figure 2.6. Two-way interaction between participants' group-based discrimination in collective identification responses and targets' group membership 98
Figure 2.7. Intra- and intergroup coalitional preferences (Experiment 2) 99
Figure 2.8. Basic anatomy of the eye 112
Figure 2.9. Assignment of fixations to areas of interest 125
Figure 2.10. Changes in pupil size during the presentation of facial expressions 134
Figure 3.1. Perceptions of emotion-perception and empathic norms (Experiment 4) 166
Figure 3.2. Perceived emotion-perception and empathic norms in relation to intergroup prejudice perceptions (Experiment 4) 170
Figure 3.3. Empathy and perceptions of empathic norms in relation to prejudices and prejudice perceptions (Experiment 4) 172
Figure 3.4. Perceptions of empathic norms (Experiment 5) 183
Figure 3.5. Perceptions of empathic norms in relation to intergroup prejudice perceptions (Experiment 5) 188
Figure 3.6. Empathy and perceptions of empathic norms in relation to prejudices and prejudice perceptions (Experiment 5) 190
Figure 4.1. Proposed mediation model from (perceived) outgroup intergroup empathic norms to outgroup-oriented empathy 212
Figure 4.2. Empathic feedback and an example of a main task trial 216
Figure 4.3. Learning phase of the empathy learning task 241
Figure 4.4. The influence of outgroup intergroup empathic norms on participants' intergroup empathy 249
Figure 4.5. The influence of the most recent outgroup intergroup empathic feedback on participants' intergroup empathy 251
Figure 4.6. Actual versus perceived intra- and intergroup empathic norms 260
Figure 4.7. Confusion matrices quantifying the probabilities that each model is the best fit to the data generated from each model 261
Figure 5.1. Overview of how ingroup and outgroup intergroup empathic norms might signal outgroup members' deservingness of empathy 307
Figure A3.4.1. Bootstrapped confidence intervals of relevant edge-weights for the estimated networks of prejudice and empathic norm perceptions 418
Figure A4.3.1. Actual versus simulated perceptions of intergroup empathic norms 429

LIST OF TABLES

Table 1.1. Matrix of perceiver's and empathic target's emotions to determine the nature of empathy and counter-empathy 31
Table 2.1. Means and standard deviations used in the simulation-based power analysis 70
Table 2.2. Descriptive statistics of included self-report questionnaires and convincingness questions (Experiment 1) 91
Table 2.3. Descriptive statistics of included self-report questionnaires and convincingness questions (Experiment 2) 101
Table A3.1.1. List of Croatian names used in Experiment 4 402
Table A3.1.2. List of Serbian names used in Experiment 4 403
Table A3.3.1. Model building and comparison for the average ingroup empathiser condition (Experiment 4) 406
Table A3.3.2. Model building and comparison for the average outgroup empathiser condition (Experiment 4) 408
Table A3.3.3. Model building and comparison for the average ingroup empathiser condition – Empathy responses (Experiment 5) 410
Table A3.3.4. Model building and comparison for the average outgroup empathiser condition – Empathy responses (Experiment 5) 412
Table A3.3.5. Model building and comparison for the average ingroup empathiser condition – Emotion perception responses (Experiment 5) 414
Table A3.3.6. Model building and comparison for the average outgroup empathiser condition – Emotion perception responses (Experiment 5) 416
Table A3.5.1. Comparison of acontextualised intergroup empathy bias participants expected from ingroup and outgroup members across experiments 419
Table A4.1.1. Overview of the fitted models 423

ABSTRACT

Judgments of others' emotions and empathy in response to these emotions are shaped by the characteristics of the expresser, including which groups they belong to. To date, most research has taken one of the two approaches to examine the influence of group membership on emotion judgments and empathy. First, a lot of past research has exposed participants to members of real-world group categories (e.g., racial, gender, or political groups) and tracked whether these categories influence how participants judge and respond to emotions. This approach has mostly refrained from generalising findings to unexamined intergroup contexts and from providing insights into group processes more broadly. Second, an equal amount of research effort has gone into exposing participants to members of novel group categories associated with arbitrary labels and tracking whether these categories influence how participants judge and respond to emotions. This approach has mostly interpreted findings in terms of "ingroups" and "outgroups", thus implying that all ingroups and outgroups should be expected to exert similar influence. Both of these approaches make it difficult to identify meaningful contextual variability across intergroup contexts and to systematise how *abstract features* of particular ingroups and outgroups, their relations, and the environments in which they operate might affect emotion judgments and empathy. For instance, should groups be expected to influence emotion judgments and empathy regardless of the specific relations between them (e.g., whether the relations are cooperative or competitive) or do intra- and intergroup relations meaningfully shape these psychological processes? To offer a more precise answer to the question of *when* (i.e., *in which intergroup contexts*) intergroup biases in emotion judgments and empathy are likely to occur, the present thesis examines the influence of *group relations* alongside *group category membership* on empathy and judgments of facial expressions. Chapter 1 reviews relevant literature, introduces the analytic lens (*feature-oriented thinking about groups*) and outlines the seven experiments comprising

this thesis. Chapter 2 investigates how group category membership and coalitional relations (i.e., competition, cooperation) influence judgments of facial expressions of emotion and whether there are any attentional and physiological markers associated with processing of social information in intergroup contexts. Whereas Chapter 2 focuses on judgments of facial expressions, Chapter 3 investigates people's subjective perceptions of social norms surrounding intergroup empathy (i.e., *intergroup empathic norms*) and how these perceptions vary as a function of how relevant group relations are perceived. Chapter 4 expands on Chapter 3 to examine the causal influence of intergroup empathic norms on one's own intergroup empathy. Specifically, Chapter 4 explores the mechanisms through which intergroup empathic norms achieve their influence, the preferences people display when faced with ingroup and outgroup norms that demand misaligned emotional reactions (intergroup apathy versus intergroup empathy) and the ways in which observable emotions can serve as cues of the underlying relations between groups. Finally, Chapter 5 discusses the theoretical and practical implications of this body of work and outlines novel directions for research looking at the interplay between socio-affective and group processes. Overall, this thesis shows that intergroup biases in emotion judgments and empathy depend not only on the general distinction between ingroups and outgroups but also on the specific relations between the ingroups and outgroups under consideration. This thesis also highlights the need for features of groups, group relations and group environments to be more meaningfully incorporated into social psychological research and theorising.

CHAPTER 1: INTRODUCTION

General overview of the chapter

Undoubtedly, groups represent a fundamental organising principle of our social lives and are recognised as such by most. When asked to identify groups in society, people easily generate a lengthy list that tends to vary depending on their socio-cultural background. For instance, "white people", "Democrats" and "black people" are spontaneously named by most in the United States as meaningful social groups, while "students", "children" and "(un)employed" reach the top of this list in Germany (Koch et al., 2016). Driven at least in part by early Social Identity Perspective, encompassing both Social Identity Theory (Tajfel, 1978; Tajfel & Turner, 1979) and Self-Categorisation Theory (Turner & Oakes, 1986; Turner et al., 1987), and by the highly influential minimal group paradigm, the dominant account of "groups" in social psychology and neuroscience closely tracks people's lay intuitions by treating "groups" as "group categories" (Cikara, 2021). This conceptualisation has inspired decades of research examining the influence of group categories on psychological processes and behaviours by either (a) indirectly or directly exposing participants to members of real-world group categories (i.e., racial, political, and religious groups), or (b) by indirectly or directly exposing participants to members of novel group categories associated with arbitrary labels (i.e., overestimators versus underestimators, Klee versus Kandinsky lovers). In this chapter, I first describe the view of groups as group categories and discuss how this view has shaped psychological research and theorising about groups and their influence on emotion judgments and empathy, which are the two psychological processes examined in this thesis. I then review limitations of this view, which include limited generalisability of findings across different intergroup contexts and across different time points within the same intergroup context and continue to emphasise the utility of adopting feature-oriented thinking about groups. Adopting feature-oriented thinking about groups alongside category-oriented

thinking about groups should motivate the systematic study of *features* of groups, group relations and group environments, therefore allowing for broader predictions regarding *what kind of groups and intergroup environments* are most likely to trigger group-based biases. Finally, I highlight existing social psychological research that already adopts feature-oriented thinking about groups and outline how this view is reflected in the present work.

Groups as group categories

Imagine leaving the house in the morning and catching the bus to work. During your commute, you can effortlessly recognise members of different group categories – *high school students* wearing their school uniforms, a *Conservative Party supporter* reading The Daily Telegraph, a *metalhead* wearing an Iron Maiden t-shirt or an *environmentalist* carrying a reusable coffee cup with a sustainability message. Even though you have never encountered these specific individuals, your ability to detect, represent, and reason about conceptually rich group categories allows you to infer these individuals' likely thoughts, beliefs, and actions. However, in contrast to categorising fruit as "apples" and "bananas", categorising people into group categories is an egocentric process, as we distinguish those with whom we share group (category) membership (so-called *ingroup members*) from those we do not (so-called *outgroup members*), which has vast perceptual, motivational, and behavioural consequences (Bodenhausen et al., 2012; Liberman et al., 2017). For instance, if you are a Conservative Party supporter yourself, you will likely be more eager to engage in a conversation with the person reading The Daily Telegraph than if you are a Labour Party supporter. This simple example provides a good starting point to discuss the dominant account of groups in social psychology, best represented by the Social Identity Perspective, encompassing both Social Identity Theory (SIT, Tajfel & Turner, 1979) and Self-Categorisation Theory (SCT, Turner & Oakes, 1986). Thus, in the next section, I provide a brief overview of the Social Identity Perspective, its empirical background, and implications. Finally, as SIT and SCT are highly intertwined and rely on the same assumptions and methods (Reimer et al., 2020), I will not present these theories in the order they were developed, nor will I focus on clarifying their similarities and differences. Instead, the next section will start by defining the core processes introduced by the Social Identity Perspective – *social categorisation*, *self-categorisation*, and

social identification – and will then relate these processes to intergroup phenomena, particularly the ones reflected in the results of the minimal group paradigm experiments.

Brief overview of the Social Identity Perspective

Social categorisation. As was illustrated in the initial bus commute example, people readily sort others into group categories, noting prominent characteristics such as age, race, and gender. On the one hand, this *social categorisation* indisputably has functional value. In particular, identifying group categories allows us to systematise otherwise overwhelming knowledge about human qualities and relations and navigate complex social environments by flexibly applying group-level knowledge to individual group members we encounter (Gaertner et al., 2010; Kunda & Spencer, 2003; Liberman et al., 2017; see also Cooley & Payne, 2019). For example, we might know that metalheads are commonly considered to be rebellious, so we might automatically assume that the specific metalhead on our bus is also rebellious. On the other hand, social categorisation's functional value comes at a cost as it profoundly changes how we process information about others, prioritising group prototypes and boundaries over the continuity and fluidity of human attributes (Leonardelli & Toh, 2015; Reimer et al., 2022). For example, we might not pay as much attention to how individual metalheads differ as we automatically apply the same prototypical representation to everyone we categorise as belonging to this group. Similarly, we might fail to recognise that many prominent characteristics, from music taste to sexual orientation, exist on a spectrum rather than discrete categories (e.g., metalhead versus pop music lover, gay/lesbian versus straight, Bjornsdottir et al., 2021; Ho et al., 2011).

Self-categorisation. Social categorisation is not reserved for others but is similarly applied to oneself. This self-oriented social categorisation or *self-categorisation* (Reimer et al., 2020) refers to the cognitive placing of oneself and others into the same group category ("ingroup members" or "us") in contrast to other salient group categories ("outgroup

members" or "them"). However, just because people are capable to categorise themselves into group categories does not mean that they always do and correspondingly, that their sense of self is exclusively derived from group belonging. On this note, SCT distinguishes between *personal identity*, part of one's self-concept derived from one's idiosyncratic characteristics, and *social identity*, part of one's self concept derived from one's belonging to social groups (Turner et al., 1987)[1]. To illustrate, in one situation, I may think of myself as *Jelka Stojanov* with my unique combination of traits and habits (e.g., *nature lover and a fan of stand-up comedy*), but in a different situation, I may think of myself as a member of a group category (e.g., *Oxford DPhil student*) with characteristics shared with other members of the same group category (e.g., *going to formal dinners and completing the Transfer of Status*).

Whether a person defines themselves and others in terms of their personal identities or one of multiple social identities shapes how they relate and interact with others. According to Tajfel (1981), social situations and corresponding social behaviours vary along an interpersonal-intergroup continuum. At the *interpersonal* extreme, social behaviour is determined by individual characteristics and interpersonal relations (e.g., an interaction between two friends or colleagues). At the *intergroup* extreme, social behaviour is determined solely by group category memberships (e.g., an interaction between members of the Oxford and Cambridge rowing teams on the day of the race). However, the same social encounter can slide between interpersonal and intergroup extremes – for instance, an interaction between two colleagues from different countries can start off as purely interpersonal but respective national identities can get increasingly salient over time, pushing this interaction closer to the intergroup extreme. When group categories (e.g., national

[1] In their seminal writings, Turner and colleagues (1987) also described a third, most inclusive level of self-categorisation: the self as a human being (*human identity*). However, this superordinate identity is less relevant for most intra- and intergroup processes and is rarely discussed in social psychology.

identities) are recognised, individuals may start to perceive themselves and others as group representatives (e.g., *Serbian person*) rather than unique individuals (e.g., *Jelka Stojanov*) – a process termed *depersonalisation* (see Hogg et al., 1995; Stets & Burke, 2000). However, given that the number of group categories one can place oneself and others into is endless, one question remains open: what determines which group categories become psychologically activated in a particular situation?

Which group category becomes salient depends on the "interaction between the *relative accessibility* of the [group] category within the perceiver's repertoire and the *fit* between stimulus input and stored [group] category specifications" (Oakes et al., 1991, p. 126). Relative accessibility can be further broken down into *situational accessibility*, which refers to the group category's accessibility in the immediate social context, and *chronic accessibility*, which refers to the person's general tendency to perceive a particular group category as important across a range of social contexts (Blanz, 1999; see also Bruner, 1957). For instance, going to a friends' house to watch Eurovision and seeing that they have decorated their house with flags corresponding to guests' countries of origin illustrates the situational accessibility of one's national identity, while frequently having one's national identity at the forefront of one's mind in the absence of external cues illustrates its chronic accessibility. Furthermore, two aspects of group category fit can be distinguished – *comparative* and *normative* (Oakes et al., 1991). Comparative fit refers to categorising oneself and others in a way that minimises the perceived differences between members within one's ingroup and maximises the perceived differences between one's ingroup and a comparative outgroup. However, it is not sufficient for group categories to minimise intragroup differences and maximise intergroup differences (i.e., comparative fit), group members' characteristics and behaviours should also align with relevant stereotypes about their respective group category (i.e., normative fit). To illustrate, imagine that you are

attending a wedding where all male guests are wearing suits and all female guests are wearing dresses. In this situation, recognising "men" and "women" as relevant group categories will not only result in sharp intergroup differences and only slight intragroup differences (i.e., comparative fit), but will also conform to the expected norm of how men and women usually dress for a wedding (i.e., normative fit).

Social identification. Tajfel (1978, p. 63) defined social identity as "[a] part of an individual's self-concept which derives from their knowledge that they belong to certain social groups [i.e., self-categorisation] together with the value and emotional significance attached to that membership". On its own, this definition suggests that self-categorisation is only one part of social identification. Specifically, self-categorisation is commonly seen as the first step or a single component of social identification (Ashmore et al., 2004; Ellemers et al., 1999; Jackson, 2002), with *social identification* referring to the extent a person views their group category membership as "meaningful, desired, and important" (Reimer et al., 2020, p. 5) which can trigger a variety of preferences for "us" over "them" (so-called *intergroup or ingroup biases*). While it is no major claim that social categorisation[2] is *necessary* for intergroup biases – people need to recognise themselves and others as members of social groups for intergroup biases to occur – SIT went a step further by proposing that social categorisation is not only necessary, but also *sufficient* to elicit intergroup biases (Diehl, 1990). But how does one isolate social categorisation from all possible confounding

[2] To adhere to the terminology of SCT more closely, thus far, I primarily used the term *self-categorisation* to refer to the placement of oneself into some group categories (ingroups) in contrast to other available group categories (outgroups). However, since I am moving into the realm of SIT which made no reference to self-categorisation but only social categorisation, from this point onwards, I will primarily use the term *social categorisation* to refer to the division of the social world into different group categories. In essence, the term social categorisation is synonymous with the term self-categorisation, with both terms describing an inherently egocentric process where the perceiver not only identifies group categories but always places oneself into one of them.

variables (e.g., differences in group members' appearance, associated stereotypes, or our prior experience with them)?

To achieve this, Tajfel and colleagues (1971) introduced the minimal group paradigm (MGP). Their MGP experiments started by assigning participants to "minimal" groups using trivial criteria (e.g., abstract art preferences). After being assigned to minimal groups, participants were asked to anonymously allocate resources to ingroup members and outgroup members, with their choices having no effect on their own outcomes. Even though minimal groups were created on the spot and participants never actually got to interact with other members of these groups, they still favoured minimal ingroup members over minimal outgroup members in the way they allocated resources. Since these early experiments, hundreds of studies have used MGP and found that minimal groups not only affect how people distribute resources, but also their willingness to cooperate (Goette et al., 2006; Yamagishi et al., 2008), mind attribution (Hackel et al., 2014), face memory (Bernstein et al., 2007; Ng et al., 2016; Van Bavel et al., 2012) emotion judgments (Lazerus et al., 2016) and empathy (Masten et al., 2010; Montalan et al., 2012).

Originally, Tajfel and colleagues (1971) interpreted people's tendency to discriminate in their allocation of resources as a product of alignment with a generic norm of group behaviour which dictates that people should behave differently towards ingroups and outgroups. However, the interpretation that rose to prominence was offered by Tajfel a year later (1972a, 1972b) under what would later become known as "positive distinctiveness theory". This interpretation can be divided into multiple components: (a) social categorisation, (b) social identity, (c) social comparison and (d) positive distinctiveness (Turner & Reynolds, 2010). First, people recognise the "us versus them" distinction introduced by MGP, and this *social categorisation* fundamentally changes the way people think of themselves and others. Specifically, people start seeing the world through the lens of

their *social identities* rather than their personal identities, thinking of themselves and others as group members rather than unique individuals. Once social identities take precedence over personal identities, people feel motivated to achieve favourable *social comparisons* in order to maintain or enhance the positivity of their own social identity, and consequentially, boost their self-esteem. Thus, ingroup-favouring allocation responses were interpreted as a pathway to achieve favourable social comparisons (i.e., *positive distinctiveness*) of one's ingroup in comparison to available outgroups, allowing people to reap the self-esteem benefits from their group belonging (Tajfel, 1972b; Turner & Oakes, 1986). However, more recent accounts have deemed this interpretation unlikely as "individuals would need to engage in relatively active fabrications if they are to meet this [self-esteem] goal [in the context of minimal groups]" (Dunham, 2018, p. 787), and proposed an alternative interpretation – it is because people evaluate themselves positively that they favour the minimal groups they are assigned to. In other words, people do not favour minimal ingroup members to boost positive evaluations of the self, but they do so *because* they positively evaluate themselves to begin with (Dunham, 2018). Another important distinction between early and more recent interpretations of MGP effects should be noted. While early interpretations saw "ingroup love" and "outgroup hate" as two sides of the same coin (Tajfel & Turner, 1979; Turner et al., 1987), more recent interpretations distinguished the two (Brewer, 1999) and contended that ingroup-favouring biases produced by MGP reflect favouritism towards one's ingroup (i.e., ingroup love), not antagonism towards outgroups (i.e., outgroup hate, Molenberghs et al., 2014; Molenberghs & Louis, 2018; Yamagishi & Mifune, 2009).

Summary

This section has provided an overview of the so-called Social Identity Perspective (Hogg et al., 2004; Hogg & Smith, 2007), encompassing both Social Identity Theory (SIT, Tajfel, 1978; Tajfel & Turner, 1979) and Self-Categorisation Theory (SCT, Turner & Oakes,

1986; Turner et al., 1987). In brief, SCT primarily emphasised that people are neither only individuals nor only group members, but that they view themselves and others both in terms of personal characteristics (*personal identities*) and group category memberships (*social identities*). Furthermore, SIT sought to explain intergroup biases as a mechanism to achieve positive social identity. Despite some conceptual ambiguities (Gillespie et al., 2012; Rabbie & Horwitz, 1988; Rabbie et al., 1989; Schiffmann & Wicklund, 1992), both theories – in combination and in isolation – remain at the heart of social psychological theorising and are justifiably credited for highlighting the profound importance of social identity processes and for showing that intergroup biases can in fact occur in the absence of factors previously considered necessary (e.g., competition for limited resources – Sherif et al., 1961). What is discussed significantly less, however, is how the Social Identity Perspective was applied in vast areas of social/cognitive psychology (e.g., social cognition research) and social neuroscience, influencing the conceptualisation and research of "groups" as "group categories" for decades to come (Cikara, 2021).

Complementing the study of group categories with the study of group features

Building on the concept of social categorisation and on the characteristics of the minimal group paradigm, most social cognition and social neuroscience research has studied "groups" as [self-contained, detectable, and labelable] "group categories" (*category-oriented thinking about groups*[3], Cikara, 2021), rarely seeking to unpack relevant features associated with these group categories and the broader social environment in which they operate (e.g., facial morphology, power relations, class, norms, see Sen & Wasow, 2016). The study of group categories (e.g., race, gender, sexual orientation, and their intersection) as such obviously matters for both theoretical inquiry and real-world applications (e.g., Devine et al.,

[3] Terms *category-oriented thinking* and *feature-oriented thinking about groups* are introduced in the current thesis.

2012) and was particularly crucial for establishing the basic layout of intergroup phenomena in the same way brain-mapping studies were crucial to obtain the working model of the brain (see Amodio, 2008). However, just as brain-mapping studies alone cannot get us far in the pursuit of understanding complex collections of functionally and/or anatomically connected brain regions that are involved in social behaviour, the study of group categories in theoretical isolation from their constituent features cannot get us far in the pursuit of understanding *what kind of groups and intergroup environments* are likely to influence psychological processes and behaviours. Therefore, in this thesis, I argue that to better understand how group processes intersect with socio-affective processes like emotion judgments and empathy, the study of group categories should be complemented with the study of group features.

Before elaborating further, it should be noted that category-oriented thinking about groups, widespread in the fields of social cognition and social neuroscience, runs somewhat counter to the contextually rich treatment of real-world groups under the Social Identity Perspective and to much of the empirical work inspired by this perspective. In particular, Social Identity Perspective often emphasised the power of social context to explain real-world intergroup behaviour (Hogg et al., 1995; Huddy, 2002) and social psychological research drawing on this perspective was thus often explicit in trying to link contextual features (e.g., group salience, group status, permeability of group boundaries) to the occurrence of various intergroup biases (e.g., in resource allocation or evaluation; Giannakakis & Fritsche, 2011; Jackson et al., 1996; Mullen et al., 1992). Nevertheless, this work has had a relatively weak impact on the fields of social cognition and social neuroscience which have tended to adopt more reductionist, category-oriented thinking about groups.

I will now start with an example of how the study of group features could lead to more insightful findings than the study of group categories in the domain of emotion recognition. Imagine a familiar scenario: you are reading two published studies where one documented a certain effect of interest among groups A and B (for example, it documented that people recognise ingroup members' emotions better than outgroup members' emotions, Elfenbein & Ambady, 2002), and the other one failed to find the same effect among groups C and D (or found the exact opposite). After scratching your head for a bit, you notice one *feature* which the examined intergroup contexts A-B and C-D differed on – for example, the study in the intergroup context A-B involved majority group members judging both minority and majority members' facial expressions while the study in the intergroup context C-D included minority group members judging both minority and majority members' facial expressions. By shifting your analytic lens away from social identities (i.e., membership in group categories A, B, C and D), you were able to identify a potentially relevant *abstract feature* of groups and group relations (i.e., majority-minority status) that might underlie the observed difference in findings.

In response to this realisation, you decide to design an experiment where you manipulate majority-minority group status and track how this factor influences emotion recognition. You also decide to team up with other researchers and recruit participants from majority-minority groups across multiple countries. If you indeed observe that majority group members are more accurate at recognising majority than minority members' emotions while minority group members do not display any difference in accuracy, your research would have provided evidence for an *abstract group feature* that can affect emotion recognition, invariant to the majority-minority groups in question. This would place your research between two extremes: (a) an extreme where group effects are interpreted exclusively in terms of sampled group categories and their unique relations (see examples of groups in Cyprus and Northern

Ireland in Psaltis & Cakal, 2016, and Ferguson & McKeown Jones, 2016, respectively) and are not generalised to derive any insights into group processes more broadly, and (b) an extreme where group effects are indiscriminately interpreted in terms of "ingroups" and "outgroups" without recognising that there might be meaningful contextual variability to explore. This is not to say that I argue against the value of exploring the particulars of specific group categories and intergroup relations or against the existence of general social categorisation principles. I simply argue that the middle research layer which prioritises features which transcend sampled group categories deserves more attention in the literature.

Furthermore, even minimal groups, which have long been considered devoid of rich meaning (Dunham, 2018, see also Hong & Ratner, 2020), usually have certain experimentally induced or assumed features (e.g., perceived similarity of and/or interdependence between group members, Diehl, 1988; Diehl, 1990; Durrheim et al., 2016; Locksley et al., 1980; Platow et al., 2012). However, in a research context where the positive valuation of social identities remains the most frequently evoked explanation of minimal group effects (Dunham et al., 2011; Hargreaves Heap & Zizzo, 2009; Otten & Epstude, 2006), the contribution of features such as perceived similarity and interdependence to intergroup biases in psychological processes and behaviours is mostly left underdiscussed and underappreciated (see Gaertner & Insko, 2000; Pinter & Greenwald, 2011). The greater recognition of these features in the context of minimal groups is necessary given that perceived similarity and interdependence have on their own been shown to produce effects similar to those of minimal group membership (Kawakami et al., 2021; Plötner et al., 2015; Sparks et al., 2017; Valdesolo & DeSteno, 2011). This raises the question of whether minimal group effects can be re-cast as by-products of these features. Although I believe that psychological research would benefit from more widespread recognition of *features* behind minimal group categories and I try to model this belief in the current thesis, I do not think

that the debate should be framed in a "either social identities or features" manner (e.g., intergroup biases either stem from positively valued minimal ingroup identity or from anticipated interdependence with minimal ingroup members) as these views are perfectly reconcilable. Specifically, people might be sensitive to perceived similarity and interdependence with others and this sensitivity, coupled with the assumption that minimal ingroup members are similar to and interdependent with one another, might contribute to social identification with minimal ingroups and occurrence of intergroup biases in MGP experiments (see Pietraszewski, 2013).

In summary, while *category-oriented thinking about groups* produces research that largely ignores group features or primarily looks at them *after* a certain effect has been observed in a certain group category or intergroup context (e.g., "people recognise racial ingroup members' emotions better than racial outgroup members' emotions, why might that be?"), *feature-oriented thinking about groups* puts features ahead of group categories. As such, feature-oriented thinking about groups produces research that seeks to generate broader inferences about groups and group processes, so that even if our social world is turned upside down and the group categories (e.g., nationalities, religions) we are used to no longer apply, we can still make predictions about how groups will form as well as how group dynamics will evolve and influence psychological processes and behaviours.

Finally, this thesis investigates the influence of multiple features of group relations alongside group category membership on judgments of facial expressions of emotion and empathy. For this reason, I will now describe existing research which has examined how groups affect judgments of facial expressions of emotion and responses of empathy, particularly highlighting theorising which has built on group-category-centred evidence to produce feature-oriented interpretations that transcend the specific group categories that were studied. Then, I will elaborate on the argument that the study of specific group categories

should be complemented with the study of features of groups, group relations and group environments in order to achieve predictions that move beyond the here and now and acknowledge that group categories and dynamics are anything but static and inflexible.

The influence of groups on judgments of facial expressions of emotion

Our judgments of others' facial expressions of emotion are shaped significantly by the characteristics of the expresser (Hess et al., 2009), including their group category membership. Recognising that a person is experiencing a specific emotion might intuitively seem like a straightforward product of "reading" visible cues off the person's face, but this process is better understood as an outcome of a dynamic interaction between bottom-up (e.g., spatial and temporal configuration of the facial expression) and top-down (e.g., pre-existing beliefs about the expresser's group) factors (Barrett, 2017; Bijlstra et al., 2010; Brooks & Freeman, 2018; Freeman & Ambady, 2011; Freeman & Johnson, 2016). Given that groups have been at the source of both humanity's greatest successes and humanity's greatest atrocities and given just how consequential intergroup misunderstandings in non-verbal communication can be (e.g., Correll et al., 2002; Correll et al., 2011; Unkelbach et al., 2008), it is understandable that a considerable body of research has focused in particular on the influence of groups on judgments of facial expressions of emotion.

Categorical emotion recognition

Most research to date has examined emotion recognition accuracy – whether the emotion category identified from the facial expression (e.g., anger) matches the emotion category the expresser was trying to convey – and has consistently documented an *ingroup advantage*, whereby people are more accurate at recognising facial expressions of emotion produced by members of their own cultural group versus members of different cultural groups (Dailey et al., 2010; Elfenbein, 2013; Elfenbein & Ambady, 2002; Elfenbein &

Ambady, 2003; Kang & Lau, 2013; Kommattam et al., 2017; van der Schalk et al., 2011; Wickline et al., 2009). Furthermore, people tend to report feeling more confident when recognising cultural ingroup members' facial expressions of emotion in comparison to cultural outgroup members' expressions (Beaupré & Hess, 2006). Whereas most studies indeed suggest better emotion recognition of own-group facial expressions of emotion, others have found no such effects (Lee et al., 2005; Matsumoto et al., 2009; Tracy & Robins, 2008) or have found the exact opposite, better emotion recognition of other-group facial expressions of emotion (*outgroup advantage,* Elfenbein & Ambady, 2002; Gitter et al., 1972). As illustrated in the previous section, focusing on group category membership as a unit of explanation makes these results difficult to reconcile – why would sharing group category membership with someone improve emotion recognition on one occasion and hinder it on another? By contrast, taking a feature-oriented approach, systematically outlining the characteristics of particular ingroups and outgroups, their relations, and the environments in which they operate and discussing how these characteristics might affect emotion recognition, could help us see how these previously irreconcilable findings are actually compatible. Therefore, what is it about ingroups and outgroups that can give rise to *ingroup advantage* on one end, and *outgroup advantage* on the other?

Dialect theory (Elfenbein et al., 2007) provides the dominant and the most well-supported interpretation of ingroup advantage in emotion recognition. According to dialect theory, facial expressions of emotion have some universal characteristics thus providing a common language for communication, but cultural groups systematically differ in their "expression dialects" or cultural styles of expressing emotions (Cordaro et al., 2018; Elfenbein et al., 2007). Consequently, the further the expresser's expression dialect is from the perceiver's expression dialect, the more unfamiliar the perceiver will be with the

expresser's facial expressions of emotion, making these expressions harder to recognise and resulting in lower accuracy and lower confidence in the perceiver.

Even though dialect theory has not directly addressed the *outgroup advantage* in emotion recognition, it can be reasonably extended to offer an interpretation of this phenomenon too. The first thing to note in this regard is that the outgroup advantage was documented in the context of minority and majority groups within the same nation. Specifically, members of minority groups (i.e., African Americans) were more accurate at recognising outgroup majority members' (i.e., White Americans') facial expressions of emotion than they were at recognising their own ingroup minority members' expressions when minority and majority groups shared a nation (Gitter et al., 1972). Given that minority participants were often undergraduate students at a majority-dominated university (Gitter et al., 1972), it is possible that their cultural expression dialect was more similar to that of majority members than of other minority members, resulting in superior accuracy in recognising majority members' facial expressions of emotion. While it is intuitive to understand that one member of a minority group might fail to understand another if they do not speak the minority language fluently, it is easy to miss the fact that the same principle could apply in the domain of facial expressions of emotion – if a member of a minority group has spent more time interacting with or simply being exposed to majority members, it is plausible that they have developed an expression dialect similar to that of majority members, which facilitates their recognition of majority members' facial expressions of emotion (De Leersnyder, 2017). This extension of dialect theory has not been tested to date, and it remains for future research to address it, for example, by quantifying the amount of minority members' exposure to majority members' culture and testing whether this moderates the size of outgroup advantage in emotion recognition (Elfenbein, 2015).

Thus far, I have briefly presented dialect theory, which is a dominant explanation of accuracy-related emotion recognition biases in real-world intergroup contexts. I focused on this theory because it represents an excellent example of how focusing on *features* of groups, group relations and group environments (i.e., similarity in group members' expression dialects), rather than on group category membership *per se* (i.e., belonging to the same versus different group category), yields more productive insights. Specifically, taking a feature-oriented approach allows for predictions which move beyond specific intergroup contexts and offers a more precise answer to the question of *when* (i.e., *in which intergroup contexts*) intergroup misunderstandings in emotion communication are most likely to arise. Specifically, dialect theory has narrowed down the prediction from "people are less accurate when recognising outgroup members' than ingroup members' facial expressions of emotion" to "the less familiar people are with outgroup members' expression dialect, the less accurate they are in recognising outgroup members' facial expressions of emotion in comparison to ingroup members' facial expressions of emotion." This more specific prediction implies that there should be greater problems in non-verbal communication between, for instance, people from Sweden and Vietnam than between people from Sweden and Denmark. In other words, the intergroup effect does not only depend on the general distinction between ingroups and outgroups but also on the specific relations between the ingroups and outgroups under consideration.

However, some evidence suggests that there is more to the ingroup advantage in emotion recognition than dialect theory suggests. For instance, it has been shown that merely sharing a group category membership with the expresser is sufficient to facilitate recognition of their facial expressions of emotion (Gamond & Cattaneo, 2016; Young & Hugenberg, 2010) and increase confidence in one's judgments (Young & Wilson, 2018). For example, the most well-known experiment conducted by Young and Hugenberg (2010) showed that

participants who had been assigned to novel groups (Red or Green group) based on a bogus personality test were more accurate in recognising minimal ingroup members' facial expressions of emotion than minimal outgroup members' expressions despite the fact that the culture and the race of the expresser and the perceiver were held constant, rendering expression dialects unable to influence participants' emotion recognition. The authors offered a motivation-based explanation of these results – people are simply more motivated to attend to even minimal ingroup members than minimal outgroup members and this increased motivation leads to more configural processing (i.e., processing of the spatial layout of facial features, Maurer et al., 2002) of ingroup members' than outgroup members' facial expressions.

These results suggest that ingroup advantage in emotion recognition can arise in the absence of different expression dialects. However, the offered interpretation fails to engage in the discussion of how any (assumed) features of minimal groups and intergroup context could have contributed to ingroup advantage alongside mere ingroup/outgroup distinctions. Focusing on these features alongside group category membership to explain minimal group effects is particularly important because recent research showed that people imbue minimal groups with richer meaning than was originally thought (see Yang & Dunham, 2019). For instance, Hong and Ratner (2020) have masterfully shown that minimal group labels (i.e., "overestimator" versus "underestimator"), thus far considered inconsequential and interchangeable, can in fact carry meaning. Specifically, members of different minimal groups may be perceived as having different category-related qualities (e.g., overestimators are more readily seen as generous while underestimators are more readily seen as timid). In some cases, the effect of specific minimal group labels can even be greater than the general intergroup effect. Furthermore, van Veelen and colleagues (2013) have provided evidence that when the basis for a group's existence is not well defined, as is the case with most

minimal groups, people project their personal characteristics onto the ingroup as a whole, which results in greater perceived intragroup similarity, prompting the question of whether "us" and "them" in minimal intergroup contexts can often be translated to "similar" and "different". On this note, both Young and Hugenberg (2010) and Young and Wilson (2018), who respectively documented increased accuracy and confidence when recognising minimal ingroup members' relative to minimal outgroup members' facial expressions, used a personality test to create minimal groups. This methodological decision made it particularly easy for participants to infer intrinsic similarity with minimal ingroup members alongside intrinsic dissimilarity with minimal outgroup members. In other words, despite the personality test being meaningless and despite the researchers not providing participants with any explicit feedback on the personality traits involved, this strategy of creating minimal groups still allowed participants to conclude that minimal ingroup members have the same personality traits as them and that minimal outgroup members have different personality traits from them, whatever those personality traits might be. Although people who share the same group category membership are indeed often similar, intrinsic similarity is not a necessary condition for defining and recognising group categories (as most people who have been assigned to teams for an icebreaker activity or a group assignment can confirm). Thus, it remains an open question how big a role inferred trait-level similarity played in the specific effects documented by Young and Hugenberg (2010) and Young and Wilson (2018).

In addition to imbuing minimal groups with assumed characteristics, people can also infer actual and potential relations with others on the basis of minimal group distinctions. On this note, it has been pointed out that at least some minimal group findings can be better understood as a product of coalitional relations (e.g., cooperation, competition) than as a product of group category membership (Gaertner & Schopler, 1998; Rabbie & Horwitz, 1988; Yamagishi & Kiyonari, 2000). For example, Lazerus and colleagues (2016)

documented people's tendency to judge minimal ingroup members' facial expressions of emotion as conveying more positive emotions than minimal outgroup members' facial expressions of emotion. However, the fact that Lazerus and colleagues' experiments always initiated competitive relations between minimal groups raises the possibility that the reported positivity bias in judgments of facial expressions of emotion was a result of these co-occurring coalitional relations rather than group category membership *per se*. Experiments 1 and 2 in Chapter 2 of this thesis directly address this possibility by treating group category membership and coalitional relations as orthogonal.

These criticisms highlight the importance of researchers being transparent about any experimentally induced group characteristics and relations or any group characteristics and relations that participants can be reasonably expected to infer from the experimental set-up (e.g., inferences of reciprocity, interdependence, similarity). In summary, while features such as intrinsic similarity (Bian & Baillargeon, 2022; Wilder, 1984) and positive interdependence (Craig & Lee, 2022; Flippen et al., 1996) commonly co-occur with sharing group category membership, it is important to recognise that group categories can exist without these features at play – (a) members of groups can in fact be very different from each other, and (b) group members' outcomes can be independent or even negatively interdependent (e.g., imagine being a member of a very competitive lab where people are not keen to share research equipment[4]). Thus, it might be more productive to move away from relying on the black box of group category membership and (a) explicitly manipulate minimal groups'

[4] Critics could argue that a structurally defined group (e.g., research lab) whose members are negatively interdependent is unlikely to be a great source of social identification or in other words, is unlikely to have "value and emotional significance attached to it" (Tajfel, 1978, p. 63). The Social Identity Perspective has dealt with such examples of groups mostly in the context of identity management strategies (Hogg, 2016), but I argue that they deserve more attention across areas of social psychological research (e.g., Will these groups similarly trigger an ingroup advantage in emotion recognition? Why/why not?).

characteristics and relations in order to identify *what kind of groups* are most likely to trigger an ingroup advantage in emotion recognition or any other group-based effect (see Cikara & Van Bavel, 2014) and (b) investigate whether there are individual differences in people's intuitions about the characteristics of minimal (and real-world) groups and group relations which reliably predict how participants respond to ingroup versus outgroup members (see Kranton et al., 2020).

In this thesis, Chapter 2 follows the first suggestion by experimentally manipulating coalitional relations (i.e., competition, cooperation) alongside group membership to examine the independent and joint influence of these factors on judgments of facial expressions. Chapter 3 follows both the first and the second suggestion in the domain of intergroup empathy. Under the first suggestion, Chapter 3 manipulated features of the context (i.e., characteristics of empathic targets and their emotional experiences) to examine how people's predictions of real-world ingroup and outgroup members' empathy vary with these contextual features. Under the second suggestion, Chapter 3 tracked how people's predictions of real-world ingroup and outgroup members' empathy change depending on how they perceive relevant group relations. Finally, experiments in Chapter 4 used minimal groups as a case study to examine the influence of intergroup empathic norms (discussed more in section "The influence of groups on empathy and the role of intergroup empathic norms") and as such, did not experimentally manipulate group features. Despite this, Chapter 4 abides by the general principle of transparent reporting and discussing of implied features of (a) groups, (b) their relations, and (c) the general environment in which these groups operate.

The influence of stereotypes and evaluations on categorical emotion recognition

Up to this point, I have mostly used examples of research into the effects of cultural and minimal groups on emotion recognition accuracy and confidence to draw attention to some limitations of relying only on group category membership to explain identified effects.

However, this should not create the false impression that this is the only domain where group effects on emotion processing have been examined. Another line of work has focused on the question of how emotion recognition accuracy and speed are affected by evaluative connotations of group categories (e.g., negative evaluation of minority groups or outgroup members more broadly, Dunham, 2011) and/or stereotypical associations between group categories and discrete emotions (e.g., between Black people and anger or women and sadness, Plant et al., 2000; Durik et al., 2006). For instance, Bijlstra and colleagues (2010) proposed that the influence of valence evaluations (e.g., how positively or negatively women and men are evaluated by White Dutch participants) shows up in emotion recognition primarily when participants are asked to classify positive versus negative facial expressions – for example, expressions of anger and happiness – while the influence of stereotypical associations (e.g., how strong is White Dutch participants' association between women and sadness and between men and anger) shows up primarily when participants are asked to classify facial expressions of discrete emotions that share the same valence – for example, expressions of anger and sadness. Supporting this distinction, positive facial expressions of emotion (i.e., happiness) were recognised faster than negative facial expressions of emotion (i.e., anger, sadness) for female expressers, while no recognition speed advantage was found for male expressers, which the authors explained as a reflection of women being evaluated more positively than men (Eagly & Mladinić, 1989 as cited in Bijlstra et al., 2010). Furthermore, facial expressions of sadness were recognised faster than facial expressions of anger in female expressers while the opposite pattern was found for male expressers, which the authors explained as a reflection of stereotypical associations between women and sadness and men and anger (also see Craig & Lipp, 2018a). While evaluative connotations of group categories and stereotypical associations between group categories and discrete emotions are frequently used to develop the rationale for studies and interpret their results

(e.g., Bijlstra et al., 2010; Bijlstra et al., 2018; Bijlstra et al., 2019; Craig & Lipp, 2018b; Kleider-Offutt et al., 2018), relationships between the endorsement of these connotations and emotion recognition and speed advantages are only occasionally assessed, mostly by correlating participants' explicit or implicit stereotypes and prejudice with their emotion recognition performance (e.g., Bijlstra et al., 2014; Halberstadt et al., 2020; Hugenberg & Bodenhausen, 2003).

However, the work of Dotsch and colleagues (2013) represents an interesting exception to the standard correlational research. Specifically, after creating positively and negatively stereotyped groups, Dotsch and colleagues (2013) saw that these stereotypes directly impacted how group members' faces were represented – for instance, if participants formed an initial impression that group members are trustworthy, they expected these group members' faces to be more trustworthy-looking – providing strong evidence that stereotype associations are directly implicated in the facial representation of novel group members, and potentially in emotion recognition of novel group members' expressions. On the other hand, it has been pointed out that some biases in recognising facial expressions of real-world group members might at least partially be driven by properties of group category members' faces – for instance, the overlap between masculine facial features and prototypical facial expressions of anger (Becker et al., 2007; Hess et al., 2009; Zebrowitz et al., 2010) – in addition to any relevant stereotypes, suggesting that stereotype-congruent emotion recognition advantages might not represent a fully top-down phenomenon. Nonetheless, it remains unresolved how much group effects on emotion recognition accuracy and speed have to do with (a) group members' facial features on one hand and (b) group-emotion stereotype associations on the other (but see Murray et al., 2021).

Dimensional judgments of facial expressions of emotion

The final and currently the least prolific stream of relevant research has moved away from studying group effects on emotion processing through the lens of discrete emotion categories, and instead focused on a set of underlying affective dimensions such as valence, which refers to the positivity or negativity of emotions conveyed by facial expressions (Kuppens et al., 2013), arousal, which refers to the "excitedness" or "calmness" of emotions conveyed by facial expressions (Kuppens et al., 2013), and genuineness, which refers to how much facial expressions are a reflection of what someone is really feeling (Miller et al., 2020). In contrast to previously discussed lines of work, this research stream does not refer to accuracy, but merely compares participants' judgments of ingroup members' facial expressions to their judgments of outgroup members' facial expressions. For instance, in the domain of valence and arousal judgments, it has been shown that people tend to evaluate real-world and minimal ingroup members' facial expressions as conveying more positive emotions as well as emotions characterised by higher arousal (Lazerus et al., 2016; Kommattam et al., 2019) in comparison to outgroup members' facial expressions. In the domain of genuineness judgments, Kunstman and colleagues (2016) have, for example, shown that the more suspicious Black participants were about White people's motives to respond without prejudice, the less genuine and the more threatening they perceived White people's smiles to be. Group effects on continuous judgments of genuineness (e.g., rating facial expressions' genuineness from "not at all genuine" to "extremely genuine") as well as categorical judgments of genuineness (e.g., classifying facial expressions as "posed/polite/fake" or "genuine/sincere/authentic") have almost exclusively been examined in the context of smiles (e.g., Friesen et al., 2019) or pain (e.g., Lloyd et al., 2022), with results usually pointing to an ingroup advantage or an ingroup-favouring bias. This line of

work will, however, not be further discussed as it is beyond the scope of this thesis, which at no point examines judgments of smiles or pain.

This research stream expanded the focus in an important direction, moving away from examining group effects on emotion category recognition to examining group effects on broader dimensions of emotion judgments. Nonetheless, the current literature is scant and restricted to very few specific groups and intergroup contexts. On the one hand, Lazerus and colleagues' (2016) experiments induced only competitive intergroup environments, making it impossible to say whether competition is necessary for intergroup biases in valence and arousal judgments of facial expressions of emotion or these biases in emotion judgments also occur in the presence of alternative coalitional relations. On the other hand, other existing experiments focused exclusively on a single real-world intergroup context – for instance, Kommattam and colleagues (2019) only asked White participants to judge Arab targets' facial expressions while Kunstam and colleagues (2016) only had Black participants judge White targets' smiles – again making it impossible to disentangle the role of different factors in observed results (e.g., facial features, expression dialects, associated stereotypes). This state of the literature does not allow the generation of any broader hypotheses about *what kind of groups and intergroup contexts* are most likely to affect valence, arousal, and genuineness judgments. To mitigate this, experiments described in the first empirical chapter of this thesis (Chapter 2) remove the limitation of Lazerus and colleagues' (2016) work and address the question of how different, systematically manipulated intra- and intergroup coalitional contexts affect valence, arousal, and genuineness judgments of facial expressions with the aim to identify more generalised features of groups and group relations that affect emotion judgments, invariant to the groups in question. Another research goal of the experiments in Chapter 2 is to clarify whether different intra- and intergroup coalitional contexts are also capable of eliciting distinct visual attention strategies to facial expressions,

which could potentially explain how biases in judgments of facial expressions are established.

Summary

In this section, I have reviewed existing social psychological research which has investigated how groups influence categorical recognition and continuous judgments of facial expressions of emotion. I have used this research to briefly illustrate the limited explanatory power of group categories and I have specifically highlighted theorising which has transcended group-category-centred empirical evidence to generate feature-oriented interpretations (e.g., *dialect theory*). Finally, I have introduced the research questions addressed by experiments in Chapter 2 and outlined how I see these experiments advancing feature-oriented thinking about groups and group relations as opposed to traditional category-oriented thinking. Now, I turn to similarly reviewing and discussing research which has investigated how groups influence a psychological process closely related to emotion recognition and judgments – empathy.

The influence of groups on empathy and the role of intergroup empathic norms

One issue that arises in any discussion of empathy concerns its definition. Specifically, it has been suggested that there are "nearly as many definitions of empathy as people working on the topic" (de Vignemont & Singer, 2006, p. 435), with a recent review identifying as many as 43 distinct definitions of the concept (Cuff et al., 2016). This lack of a consistent definition and the frequent confusion of empathy with related concepts such as perspective taking and compassion has had a negative impact on both research and evidence-based interventions (Cuff et al., 2016) and has even led some researchers to advocate abandoning the term empathy altogether in favour of narrower and more operationally precise terms (Hall & Schwartz, 2019). Therefore, to avoid the same conceptual confusion in

this thesis, I first review and critically assess the two dominant accounts of empathy that are most often discussed and explicitly or implicitly adopted in research examining empathy in intergroup contexts. Then I move onto reviewing interdisciplinary research which has documented failures of empathy at the intergroup level and their consequences and investigated the utility of different interventions, particularly those that target social norms (i.e., *empathic norms*) to address these empathic failures. When highlighting the most influential research on the topic, I strive to maintain complete transparency regarding precisely how empathy and empathic norms have been defined and especially operationalised. Furthermore, as in the section on facial expressions of emotion, I provide examples of how category-oriented thinking about groups has shaped research and theorising in the domain of empathy and suggest what the next phase of psychological research should do to address the limitations of this category-oriented thinking. Finally, I clarify how empathy and empathic norms are conceptualised in the current work and introduce the research questions addressed by experiments in the remaining two empirical chapters – Chapters 3 and 4.

Defining empathy

Empathy as an overarching construct. Some researchers argue for a broad, multi-component view of empathy. Under this view, empathy is conceptualised as an overarching construct encompassing three dissociable, but related and often co-occurring components – cognitive, affective and motivational (Weisz & Cikara, 2021; Zaki, 2014; Zaki & Cikara, 2015). The cognitive component of empathy, known in the field as *perspective taking* (e.g., Bruneau & Saxe, 2012), *mentalising* (e.g., Arioli et al., 2021), *Theory of Mind* (e.g., Kanske et al., 2015) or *cognitive empathy* (e.g., Schnell et al., 2011), refers to people's cognitive engagement with others' emotional states, wherein they consider empathic targets' beliefs, thoughts, and emotions (Weisz & Cikara, 2021). The affective component of empathy,

known in the field as *affect sharing* (e.g., Coll et al., 2017), *experience sharing* (e.g., Jospe et al., 2020), or *affective empathy* (e.g., Edele et al., 2013), refers to people's vicarious adoption of others' emotional states, wherein they feel what they perceive the empathic target to be feeling (Weisz & Cikara, 2021). Finally, the motivational component of empathy, known in the field as *empathic concern* (e.g., Hodges et al., 2010), *sympathy* (e.g., Wispé, 1986) or *compassion* (e.g., Preckel et al., 2018), refers to a more general feeling of concern for others' welfare which prompts the desire to help (Weisz & Cikara, 2021).

Thus far, the three components have been shown to capture most of people's lay intuitions about what empathy is (Hall et al., 2021) and to co-occur frequently in everyday experiences of empathy (Depow et al., 2021). Nonetheless, there is considerable evidence that cognitive, affective, and motivational components of empathy should be treated as distinct (Weisz & Cikara, 2021). The first line of evidence for the separability of the components is psychometric. Specifically, the latent structure assessments of most frequently used measures of empathy have revealed separable factors that map onto at least two, if not all three proposed components of empathy. For instance, the Interpersonal Reactivity Index (IRI, Davis, 1983) has distinct perspective taking and empathic concern subscales which correspond to the cognitive and the motivational component of empathy respectively, the Questionnaire of Cognitive and Affective Empathy (QCAE, Reniers et al., 2011) has a two-factor structure which, as the name suggests, differentiates the cognitive from the affective component of empathy, while the factor analysis of the Jefferson Scale of Empathy (JSE, Ward et al., 2009) further supported the separability of all three components of empathy.

The second line of evidence concerns the separability of the socio-affective outcomes of the components. A recent review by Weisz and Cikara (2021) has systematised the nuanced pattern of relationships between the three components of empathy and different socio-affective outcomes across four domains – (a) prosocial behaviour, (b) occupational

well-being, (c) relationship quality, and (d) negotiation performance – highlighting that a more fine-grained examination of empathy's components is necessary to design interventions which effectively promote desired outcomes (see Zaki & Cikara, 2015, for a similar recommendation).

Finally, the third line of evidence for the separability of empathy components stems from social neuroscience, notably from a combination of neuroimaging and lesion studies. Specifically, neuroimaging research has consistently linked cognitive, affective, and motivational components of empathy to activity in different neural networks (e.g., Kanske et al., 2015; Kanske et al., 2016; Weisz & Zaki, 2018; Zaki & Ochsner, 2012), with lesion studies further corroborating this dissociation by directly relating localised brain dysfunction to deficits in different components of empathy (Shamay-Tsoory et al., 2009).

In conclusion, this account can most succinctly be described as an "empathy umbrella" (Weisz & Cikara, 2021) and can be summarised in the following definition of empathy – "empathy is a multidimensional construct which includes understanding, sharing and feeling concern for others' emotions" (Zaki & Cikara, 2015, p. 471) – where the cognitive component corresponds to "understanding", the affective component corresponds to "sharing" and the motivational component corresponds to the "feeling concern" part of the definition. While some may see this definition's comprehensiveness as a strength, it can easily be construed as a weakness. Most importantly, it is unclear what the added value of clinging to the second-order term *empathy* is, especially since most proponents of the "empathy umbrella" view in fact advocate a more fine-grained examination of individual components (e.g., Weisz & Cikara, 2021; Zaki & Cikara, 2015) to achieve more nuanced understanding and greater conceptual clarity (see also Hall & Schwartz, 2019). For this reason, in this thesis, I do not use the term empathy to refer to a complex three-component

construct, but instead adhere to a more narrow view of empathy, presented in more detail in the upcoming section.

Empathy as affect sharing. In contrast to the aforementioned three-component view, some researchers prefer a narrower definition of empathy, wherein empathy is a purely affective process and refers to the perceiver vicariously adopting emotions identified in the empathic target (Cuff et al., 2016; de Vignemont & Singer, 2006; Singer & Klimecki, 2014). One open question, however, concerns the degree to which perceiver's emotions need to match the emotions identified in the empathic target for the perceiver's reaction to be considered empathic (de Vignemont & Singer, 2006). In other words, it is currently unclear whether it is sufficient for the perceiver's emotions to only roughly match those identified in the empathic target by, for example, sharing the same valence, or whether the perceiver's and the empathic target's emotions need to match not only in valence, but also in intensity and their categorical label (e.g., anger, sadness).

To date, most research investigating empathy in intergroup contexts has steered clear of this debate and treated congruent valence between the perceiver's and the empathic target's emotions as a sufficient condition for the perceiver's reaction to be considered empathic. For instance, if the perceiver feels positively in response to empathic targets' positive emotions, they are thought to be experiencing *positive empathy* and if the perceiver feels negatively in response to empathic targets' negative emotions, they are thought to be experiencing *negative empathy* (Hudson, 2022, also see Table 1.1). Correspondingly, incongruent valence between the perceiver's and the empathic target's emotions is a sufficient condition for the perceiver's reaction to be considered counter-empathic. Based on the specific valence incongruency between the perceiver's and the empathic target's emotions, counter-empathy can be divided into two discrete emotions - *schadenfreude*, which refers to the perceiver feeling positively in response to empathic targets' negative emotions,

and *gluckschmerz*, which refers to the perceiver feeling negatively in response to empathic targets' positive emotions (Smith & van Dijk, 2018, also see Table 1.1).

Table 1.1

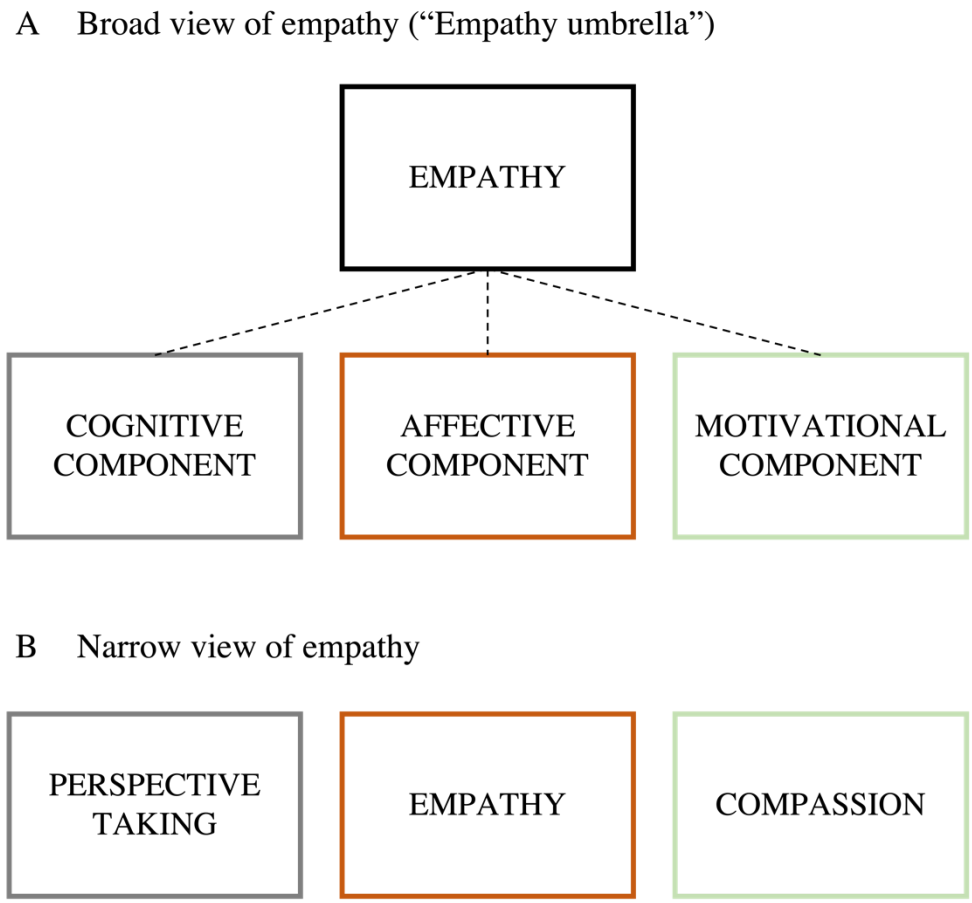
Matrix of perceiver's and empathic target's emotions to determine the nature of empathy and counter-empathy

		Perceiver's emotions	
		Positive	Negative
Emotions identified in the empathic target	Positive	Positive empathy	Gluckschmerz
	Negative	Schadenfreude	Negative empathy

In conclusion, the difference between the narrow and the broad view of empathy lies only in the question of what the term "empathy" refers to (see Figure 1.1). Specifically, empathy under the narrow view corresponds only to the affective component of empathy under the broad view, wherein the perceiver is thought to be experiencing empathy if (a) they feel emotions isomorphic to empathic target's emotions (in terms of valence), (b) these emotions are triggered either by observing or imagining empathic target's emotions, and (c) the perceiver is aware that the source of their emotions lies in another person (de Vignemont & Singer, 2006). It is important to emphasise that the line of research which adheres to the narrow definition of empathy has only altered the terminology but not the breadth of processes being studied. Specifically, this line of research has not abandoned the examination of perspective taking and compassion but treats these as stand-alone processes at the same level as empathy instead of components of empathy (see Figure 1.1).

Figure 1.1

Visualisation of the two discussed views of empathy



Notes. As indicated by the colours, perspective taking under the narrow view corresponds to the cognitive component of empathy under the broad view, empathy under the narrow view corresponds to the affective component of empathy under the broad view while compassion under the narrow view corresponds to the motivational component of empathy under the broad view. Finally, it is important to note that each of these processes – perspective taking, empathy, and compassion – has also been described using other terms (e.g., perspective taking is also referred to as mentalising or Theory of Mind, empathy is also referred to as experience sharing or affect sharing, while compassion is also referred to as empathic concern or sympathy), but for consistency, perspective taking, empathy, and compassion will be the terms used in this thesis.

This thesis aligns with the narrow view of empathy. Thus, the term "empathy" will henceforth be used to refer to the vicarious adoption of the empathic target's emotions in terms of valence – for example, if participants report feeling negatively in response to observing an empathic target suffering a negative outcome, they would be considered to have responded empathically. Similarly, the term "counter-empathy" will henceforth be used to

refer to the adoption of emotions that are opposite to what the empathic target is experiencing in terms of valence – for example, if participants report feeling positively in response to observing an empathic target suffering a negative outcome or feeling negatively in response to observing an empathic target achieving a positive outcome, they would be considered to have responded counter-empathically, or more specifically, to have responded with schadenfreude or gluckschmerz respectively. Please note that my alignment with the narrow view of empathy is only a matter of vocabulary and that across chapters, this thesis occassionaly examines and discusses perspective taking (i.e., cognitive component of empathy under the "empathy umbrella" view) and compassion (i.e., motivational component of empathy under the "empathy umbrella" view). In other words, while I fully acknowledge that perspective taking, empathy and compassion frequently co-occur (Depow et al., 2021), I treat them as stand-alone constructs rather than lower-level (cognitive, affective, and motivational) components of a higher-order construct of empathy. Finally, empathy and counter-empathy have thus far been illustrated in categorical terms for the sake of simplicity (e.g., empathiser feels "positively" or "negatively" in response to empathic target's emotions). However, I fully recognise that these emotional reactions vary in intensity (e.g., a perceiver's feelings in response to an empathic target's emotions can fall anywhere between "not positive/negative at all" and "extremely positive/negative") and therefore, I examine them as continuous in the current work.

Intergroup empathy (failures)

Empathy – vicarious sharing of other people's emotions – has long been praised for its socio-affective benefits. For instance, experiencing empathy has been suggested to facilitate interpersonal understanding and improve relationship quality (Gable et al., 2004; Randall & Schoebi, 2015; Scheffer et al., 2021) as well as to motivate prosocial behaviour (Klimecki, 2019; Lehmann et al., 2022, but see also Jordan et al., 2016). Thus, empathy can

be considered a crucial building block of cohesive social relations. However, just because people have the ability to empathise does not mean that they always do (Keysers & Gazzola, 2014).

One of the most well documented moderators of empathy concerns the empathic target's group category membership. In particular, mounting evidence indicates that people empathise more with ingroup members (e.g., members of the same racial, national, or even lab-created group) than with outgroup members (e.g., members of a different racial, national, or even lab-created group; Bruneau et al., 2015; Bruneau et al., 2017; Montalan et al., 2012; Vollberg et al., 2021). This *intergroup empathy bias* is not solely an intriguing line of social psychological inquiry, but has been shown to have profound, real-world consequences – for example, feeling apathy in response to outgroup suffering has been associated with neglect (i.e., passive harm), while the feeling of schadenfreude has been theorised to be a strong motivator of overt intergroup aggression (i.e., active harm, Cikara, 2015).

While some studies have relied on explicit self-report measures to examine the influence of group category membership on empathy (e.g., Bruneau et al., 2017) and counter-empathy (e.g., Hoogland et al., 2015), most research has sought to unpack this influence using neuroimaging measures such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) as these measures are less susceptible to social desirability effects (Cikara et al., 2011). A great deal of neuroimaging work has centred around empathy for pain, building on a well-supported finding that feeling empathy for others' pain activates similar areas of the brain as the first-hand experience of pain, primarily brain regions associated with affective as opposed to sensory aspects of pain such as the anterior cingulate cortex (ACC) and anterior insula (AI, Singer et al., 2004). This finding has been taken as evidence that shared neural representations for own and others' pain, and likely emotions in general, lie at the core of how empathy operates (Contreras-Huerta et al., 2013).

This has inspired an abundance of research investigating whether group boundaries dampen these shared neural representations.

In one empathy-for-pain fMRI study (Xu et al., 2009), Caucasian and Chinese participants watched both Caucasian and Chinese targets receiving either painful (e.g., a needle poke) or non-painful stimulation (e.g., Q-tip touch). Both groups of participants showed increased blood-oxygen-level dependent (BOLD) activity in the ACC when observing racial ingroup members relative to racial outgroup members receiving a painful stimulation, which was taken to signal a racial bias in empathic responses to others' pain. In a similar empathy-for-pain fMRI study (Hein et al., 2010), participants not only showed increased activity in the left AI when witnessing an ingroup member (i.e., fan of the same football club) as opposed to an outgroup member (i.e., fan of the rival football club) receiving a painful electric shock, but they also showed increased activity in the ventral striatum (VS), a brain region associated with reward processing and pleasure (Takahashi et al., 2009), when observing an outgroup member in pain relative to an ingroup member in pain. The VS activity recorded when participants watched outgroup members receiving painful electric shocks was negatively associated with participants' willingness to help outgroup members. Specifically, the greater VS activity participants showed, the less willing they were to receive an electric shock to reduce the intensity of outgroup member's pain stimulation.

Since these original findings, the intergroup bias in empathic neural responses to the pain of others has been further corroborated in a number of EEG and fMRI studies which have focused on multiple, usually racial, group categories and used a range of different paradigms (Azevedo et al., 2013; Cheon et al., 2011; Gutsell & Inzlicht, 2012; Krautheim et al., 2019; Riečanský et al., 2015). Note, however, that there are also several studies which have not found intergroup bias in empathic neural responses. For instance, Bruneau and colleagues (2012) showed that Arab and Jewish Israeli participants exhibited equivalent

ACC, AI, and medial prefrontal cortex[5] (mPFC) activity in response to ingroup and outgroup members' painful experiences, which did not align with participants' explicit reports of lower compassion for outgroup members' relative to ingroup members' suffering. Similarly, in Mathur and colleagues' (2010) experiment, only Black participants showed greater ACC and AI activity to ingroup relative to outgroup targets' pain, while White participants' ACC and AI responses were equivalent regardless of who the suffering target was. Discrepancies like these suggest at least two domains that require further research. First, more research is needed to understand the precise role of frequently studied brain regions (e.g., ACC, AI, mPFC) in supporting psychological processes such as empathy and compassion (Vollberg & Cikara, 2018). Second, future research should recognise that different ingroups and outgroups are not equivalent and interchangeable and embrace feature-oriented thinking about groups, group relations, and group environments to identify the general circumstances under which behavioural and neural intergroup empathy biases are most likely to occur, regardless of the specific group categories (e.g., Asian-Black-White, Arab-Jewish Israeli-South American) being studied (Cikara & Van Bavel, 2014). In other words, testing independent and interactive effects of a few theoretically relevant features of groups, group relations and group environments (e.g., groups' social status or intra- and intergroup coalitional relations) should result in more valuable predictions and help reconcile these seemingly discrepant findings.

In summary, despite its benefits and moral desirability, empathy does not always occur automatically, with failures of empathy being particularly likely when the empathic target is socially distant – for instance, when they are a member of a different social group.

[5] Medial prefrontal cortex (mPFC) is part of a wider mentalising network which also includes the temporo-parietal junction (TPJ), inferior frontal gyrus (IFG), posterior cingulate and the precuneus (Rothmaler et al., 2022). Thus far, mPFC has most consistently been shown to activate when people think about own and others' mental states, whether in terms of permanent dispositions or fleeting intentions (Frith & Frith, 2006).

Following the distinction proposed by previous research, wherein people's failure to empathise is considered to reflect either their lack of ability or their lack of propensity to empathise (Keysers & Gazzola, 2014), the question remains of whether people display intergroup empathy bias because they are *unable* to empathise with outgroup members or because they are *unwilling* to do so. While some portray empathy as a "limited mental resource" that we quickly use up for those in our immediate circles, leaving us *unable* to scale empathy outward towards those who are more socially distant (Decety & Cowell, 2014; Waytz, 2016), most researchers prioritise a motivation-based explanation of intergroup empathy bias without appealing to the idea of empathy as a limited resource (Cameron et al., 2017; Cameron et al., 2022; Weisz et al., 2021; Weisz & Zaki, 2018; Zaki, 2014). Under this view, empathy is conceptualised as a motivated phenomenon where different factors can either increase (*approach motives*) or decrease (*avoidance motives*) people's willingness to empathise (Zaki, 2014). For instance, given that empathy is often perceived as challenging and cognitively costly (Cameron et al., 2019), which in itself constitutes an avoidance motive, people might not be motivated to empathise with outgroup members' (i.e., socially distant others') emotions as much as with ingroup members' (i.e., socially proximate others') emotions (Cameron et al., 2017). From this starting point, intergroup empathy bias can be exacerbated in the presence of additional avoidance motives, thereby pushing people even further away from empathising with outgroup members' emotions. For instance, people might be even less motivated to empathise with outgroup members' emotions when they do not perceive intergroup empathy as a moral imperative or even perceive it as socially undesirable, or when they do not perceive strengthening intergroup social bonds as valuable or even perceive strengthening these social bonds as detrimental to the interests and goals of the ingroup (Zaki, 2014). By contrast, intergroup empathy bias can be mitigated in the presence of approach motives that can counter the initial empathy aversion (or compassion

aversion, Scheffer et al., 2021). For instance, people might be more motivated to empathise with outgroup members' emotions when they believe that relevant others (e.g., peers) value empathy (Tarrant et al., 2009) or when they believe that empathy does not lead to emotional exhaustion (Cameron et al., 2016).

In the next section, I will review the most frequently examined and applied interventions to encourage intergroup empathy. Following the terminology offered by Weisz and colleagues (2021), I will start with *ability/knowledge/skills-based interventions*[6], which primarily target people's *knowledge* of empathic targets' circumstances and/or their *ability* to empathise across group lines. Then I will focus on *motive-based interventions*, which, as the name suggests, explicitly target people's *motivation* to empathise with outgroup members[7]. Nevertheless, the distinction between ability/knowledge/skills-based and motive-based interventions should be treated as an illustrative summary of the abundance of available empathy-building interventions, and not as a strict and impermeable categorisation of these interventions. Specifically, at least a few available ability/knowledge/skills-based interventions likely bring about their effect by increasing people's motivation to empathise in addition to their ability to empathise and their knowledge of empathic targets' circumstances. Nevertheless, people's motivation to empathise has not usually been recognised as an explicit target of these empathy-building interventions.

[6] The authors referred to this category of interventions as "skills-based interventions". The words "ability" and "knowledge" were added in this thesis to capture the variability of these interventions more adequately.

[7] It should be noted that most empathy-building interventions to date have aimed to enhance empathy in general, not specifically intergroup empathy (e.g., Weisz et al., 2021; Weisz et al., 2022). However, these general interventions have either been shown not to be effective in increasing intergroup empathy (e.g., Weisz et al., 2021), suggesting that targeted interventions might be more appropriate, or their effects on intergroup empathy were never assessed (e.g., Klimecki et al., 2014). For these reasons, this introduction focuses only on empathy-building interventions specifically designed to reduce intergroup empathy bias.

Interventions to encourage intergroup empathy

Ability/knowledge/skills-based interventions. Empathy has long been at the core of many conflict prevention and resolution efforts. Until recently, however, empathy-building interventions have operated on the premise that people do not empathise with outgroup members because they either do not have enough information about outgroup members' hardship or they otherwise lack the knowledge and skills to effectively empathise across group lines (see Lam et al., 2011; Teding van Berkhout & Malouff, 2016). Therefore, in an attempt to fill in people's knowledge gaps and improve their ability to empathise with outgroup members, empathy-building interventions have mostly relied on one of the following four techniques: (a) instructing people to take the perspective of outgroup members and to vicariously experience their hardship, (b) educating people about empathy-relevant topics, (c) using engaging audio-visual material depicting outgroup members' struggles, and (d) delivering skills training (Davis & Begovic, 2014).

A good example of the first technique is a classic study conducted by Batson and colleagues (1997). In this study, participants were instructed to either vicariously experience what a person with AIDS (Experiment 1) or a homeless person (Experience 2) go through (high-empathy condition)[8] or to maintain an objective stance towards empathic targets' experiences (low-empathy condition)[9]. Subsequently, participants in the high-empathy condition reported more positive feelings towards people with AIDS and the homeless as a group than participants in the low-empathy condition (see also Drwecki et al., 2011; Tarrant

[8] The experimental instructions in the high-empathy condition said the following: "Imagine how [the target] who is interviewed feels about what has happened and how it has affected [their] life. Try to feel the full impact of what [the target] has been through and how [they] feel as a result".

[9] The experimental instructions in the low-empathy condition said the following: "Take an objective perspective toward what is described. Try not to get caught up in how [the target] who is interviewed feels; just remain objective and detached."

& Hadert, 2010). Even though people's motivation to empathise was not explicitly acknowledged as the target of the empathy-building interventions relying on this technique, these interventions most likely brought about their effect through multiple channels: (a) by increasing people's knowledge about important aspects of empathic targets' experiences, (b) by training people in perspective taking and empathy through explicit instruction and direct experience[10], and finally, (c) by increasing people's motivation to empathise with these empathic targets in the future.

Another intervention, designed to foster reconciliation between the Hutu and the Tutsi ethnic groups in Rwanda, relied more on the second, "educational" technique (Staub et al., 2005). Specifically, Staub and colleagues (2005) designed a nine-day intervention which, among other things, included numerous lectures and group discussions on a variety of empathy-relevant topics – genocide, trauma, victimisation, healing, own and others' socio-affective needs and how these needs are affected by traumatic experiences. In comparison to the control conditions, this intervention was found to reduce symptoms of trauma and lead to more positive feelings towards members of the other ethnic group (Staub et al., 2005). Nonetheless, it should be noted that lectures and group discussions on empathy-relevant topics were delivered in conjunction with more interactive techniques, making it difficult to assess how much lectures and group discussions independently contributed to the overall reduction of trauma symptoms and improvement of outgroup-oriented feelings (see also Paluck, 2009, for an example of how educational empathy-relevant messages have been combined with other techniques to reduce intergroup prejudice in Rwanda).

The work of Soble and colleagues (2011) can be taken as an example of embracing audio-visual media. Specifically, in this study, White participants who were assigned to the

[10] For this reason, these empathy-building interventions are sometimes termed "experiential" (Lam et al., 2011; Weisz & Zaki, 2017).

experimental condition watched a brief video depicting the experiences of two men, one Black and one White, in a variety of everyday contexts (e.g., shopping at a retail store), while White participants who were assigned to the control condition watched an equally long video depicting college students' career decisions. Participants were not explicitly instructed to empathise. Witnessing the everyday struggles of racial outgroup members led to significant increase in White participants' racial awareness, as well as in feelings of empathy and guilt (see also Paluck, 2009). That being said, it remains a possibility that the video used in the experimental condition differed from the video used in the control condition on multiple features (e.g., the experimental video could have simply made racial identities more salient than the control video or could have included more empathy-evoking content than the control video), limiting our ability to isolate the exact feature of the video (e.g., depiction of racial outgroup members' hardship) which contributed to the observed improvements. Finally, similarly to the first example, it should be noted that even though this intervention was designed with educational intent, targeting students' knowledge of racial inequality, it is likely that this intervention brought about its effect by increasing students' motivation to empathise in addition to their knowledge.

Finally, to the best of my knowledge, the "skills training" technique, which seeks to teach people to more effectively express empathy regardless of whether they are actually feeling it (Davis & Begovic, 2014, see also Weisz & Zaki, 2017), has not thus far been used to increase intergroup empathy or improve intergroup attitudes more broadly. Nonetheless, some existing interventions have aimed to train people to develop a general prosocial orientation towards others, and as such, warrant a mention under the "skills training" heading despite not targeting empathy expression *per se*. In one such intervention (i.e., loving-kindness intervention, Hutcherson et al., 2008), for example, participants attended weekly meditation classes designed to help them extend warm and caring feelings usually reserved

for very close others (i.e., family members) to all human beings. In comparison to a closely matched control condition and a waitlist control condition, loving-kindness intervention led to a decrease in implicit bias against stigmatised outgroups (Kang et al., 2014), as indicated by the Implicit Association Test (Greenwald et al., 2009). Since these original findings, the ability of meditation-based interventions to reduce implicit bias has received further empirical support (Lueke & Gibson, 2015), but more research is needed to understand how these interventions affect a variety of relevant intergroup outcomes. Finally, it is important to note that the meditation-based interventions described above can be better understood as training compassion rather than empathy (see also Klimecki et al., 2014).

The empathy-building interventions enumerated so far have primarily (but not exclusively) focused on enhancing individuals' knowledge of empathic targets' circumstances and their ability to empathise across group lines by (a) explicitly instructing people to engage in perspective taking and/or to vicariously experience outgroup members' struggles, (b) educating people about empathy-relevant topics, (c) illustrating outgroup members' struggles using audio-visual media, or (d) delivering (compassion) training programmes. However, as was previously discussed, the role of motivation in intergroup empathy has gained growing recognition in recent years, inspiring interventions which explicitly strive to encourage people to be willing to empathise with outgroup members (Weisz & Zaki, 2017). These motive-based interventions have heretofore primarily targeted two sources of empathic motivation – (a) beliefs surrounding (intergroup) empathy, and (b) social norms surrounding (intergroup) empathy.

Motive-based interventions targeting empathy-related beliefs. Psychology has a long history of demonstrating the power of beliefs; their ability to affect motivation and behaviour can hardly be overstated. Empathy-building interventions that have sought to change people's empathy-related beliefs have drawn inspiration from classical work on

implicit theories about the nature of personal attributes (Chiu et al., 1997; Dweck et al., 1995a; Dweck et al., 1995b; Hong et al., 1999). For instance, reliable individual differences have been identified in how malleable people perceive intelligence as well as a variety of other personal attributes. Thus, some people see intelligence as a fixed trait that cannot be developed (i.e., *fixed or entity theory of intelligence*) while others see intelligence as more malleable and prone to development (i.e., *malleable or incremental theory of intelligence*, Haimovitz & Dweck, 2016). These implicit theories have consequences. Notably, whether a person holds a fixed or malleable theory of intelligence significantly determines what goals they pursue (proving one's abilities versus learning) and how they react to challenges or failures (helplessness versus learning from mistakes, Yeager & Dweck, 2020). Drawing a parallel with fixed and malleable theories of intelligence, Schumann and colleagues (2014) proposed that people similarly differ in their implicit theories of empathy – with some seeing empathy as a fixed disposition and other seeing it as a malleable ability – and that the theory that people endorse determines their willingness to exert effort to feel empathy. Going beyond correlating implicit theories of empathy with levels of empathic effort, Schumann and colleagues (2014) showed that experimentally inducing a malleable theory of empathy versus a fixed theory of empathy resulted in greater empathic effort in social contexts where feeling empathy was challenging – when talking about a personally important topic with someone who holds drastically different views or when listening to an emotional story of a racial outgroup member (Schumann et al., 2014)[11].

[11] Please note that the experimental manipulations Schumann and colleagues (2014) used to induce beliefs about empathy's fixed versus malleable nature portrayed empathy as a combination of perspective taking and (affective) empathy ("the social workers taught these children about *putting themselves in other children's shoes*, *trying to see things from other children's points of view*, and *feeling what other children are feeling*"). Similarly, the measure of empathic effort also referred to a combination of perspective taking ("when I disagree with someone, *I try to understand their emotions*") and (affective) empathy ("when I disagree with someone, *I try to feel what they are feeling*").

Another study focused on people's beliefs about the utility of intergroup empathy (Santos et al., 2022). Specifically, people who were led to believe that empathy for political outgroup members is a useful resource which increases their political persuasiveness reported more positive feelings toward political outgroup members and were more open to bipartisan governance than people who were led to think about intergroup empathy as a weakness which threatens their political position (Santos et al., 2022)[12]. Importantly, the effects of these manipulated beliefs extended beyond the individual participants themselves – in comparison to people who saw intergroup empathy as a weakness, those who saw it as a useful resource more frequently communicated with political outgroup members in a way that appealed to shared goals and identities and signalled perspective taking. As a result, political outgroup members developed more positive feelings towards these individuals as well as their political group as a whole (Santos et al., 2022).

In short, studies manipulating people's beliefs about (intergroup) empathy and tracking the downstream consequences of these beliefs remain scarce, but the two publications described above have yielded valuable insights into how people's intuitions about empathy's malleability and utility can promote more empathetic and positive intergroup attitudes and interactions. Finally, in the next section, I review research which is of immediate relevance for Chapters 3 and 4 in the current thesis. Specifically, this research

[12] Please note that the experimental manipulations Santos and colleagues (2022) used to induce beliefs about empathy's low versus high utility also portrayed empathy as a combination of perspective taking ("it is important to take a step back and understand that the *excessive focus on someone else's perspective* can lead us to lose track of our own ideals") and (affective) empathy ("by *absorbing the feelings of others*, those motivated to feel empathy are less equipped to successfully navigate debates"). However, in comparison to a more balanced portrayal of empathy in Schumann et al.'s (2014) work, Santos et al.'s (2022) experimental manipulations were significantly more centred around perspective taking than (affective) empathy.

has examined yet another important source of empathic motivation – social norms or more precisely, *empathic norms*.

Motive-based interventions targeting empathic norms. Decades of work show that social norms have a wide-reaching and enduring influence on a variety of phenomena – from prejudice to pro-environmental behaviours (McDonald & Crandall, 2015). People are a reflection of their social milieu and just as they closely adhere to social norms when judging the acceptability of discrimination or expressing prejudice and hostility (Crandall et al., 2002), they also adhere to social norms when expressing empathy and engaging in prosocial behaviours (Nook et al., 2016; Tarrant et al., 2009). Notably, Nook and colleagues (2016) documented the remarkable flexibility and breadth of people's conformity to social norms by showing that people do not merely copy the specific prosocial behaviours and empathic expressions seen as normative among relevant others, but more broadly conform to relevant others' prosocial orientation across emotions and behaviours. For instance, people who had observed others' generous charity donations not only donated more generously themselves but also felt more empathy for another person (conformity jump from prosocial behaviour to empathy). Correspondingly, those who learnt that their group showed high (versus low) levels of empathy for others not only reported feeling more empathy themselves but also donated more generously to a homeless shelter (conformity jump from empathy to prosocial behaviour).

Prior to elaborating on how social norms have been harnessed to motivate intergroup empathy, it is necessary to define social norms and distinguish them from a closely related concept – conventions. While the boundary between these concepts remains debated (Hawkins et al., 2019), *conventions* are commonly defined as behavioural regularities which serve primarily coordination purposes and whose violation does not evoke disapproval or sanctions, whereas *social norms* are commonly defined as *binding* behavioural regularities

that set standards for what is correct, appropriate and desirable (Bicchieri, 2006; Lewis, 1969; Nord, 1991). The distinction between conventions and social norms, which is primarily used in the field of philosophy (Bicchieri, 2006), shares some similarities with the distinction between descriptive and injunctive social norms, which is primarily used in the field of social psychology (Cialdini, 2003). Specifically, descriptive social norms reflect how relevant others in fact feel and behave (i.e., "is"), while injunctive social norms signal what emotions and behaviours relevant others approve or disapprove of (i.e., "ought", Cialdini, 2003). In practice, regardless of whether one adopts the former or the latter teminology, these concepts are often blurred, with people spontaneously inferring "ought" from "is" (Lindström et al., 2018; Tworek & Cimpian, 2016) and thus allowing what might have started as a mere descriptive social norm (or a convention) to take on a normative force. In the current thesis, all measurements of social norm perceptions (Chapter 3) and experimental manipulations of social norms (Chapter 4) *methodologically* reflect descriptive social norms (or conventions[13]) as the normative force was never experimentally induced or implied. However, one theoretical assumption guiding this thesis is that participants viewed measured and manipulated descriptive social norms as implying an "ought". In other words, when participants adapt to experimentally manipulated social norms (Chapter 4), I am assuming that they are not superficially imitating a behavioural rule, but rather inferring the presence of normative expectations that apply to anyone in one's reference group.

Furthermore, following Nook and colleagues' (2016) terminology, in this thesis, I adopt the term (*intergroup*) *empathic norms* to refer to social norms which signal how much (intergroup) empathy is commonly felt as well as perceived as desirable in one's reference

[13] Please note that some researchers (e.g., Bicchieri, 2006) view conventions as *temporally stable* descriptive social norms, so fully equating descriptive social norms with conventions is a significant simplification in the eyes of at least some theoretical positions. However, it would take a stand-alone thesis to do this topic justice, and I acknowledge that this summary is inevitably stripped of nuance.

group. For clarity, I reserve the term *empathic norms* for social norms that showcase and prescribe *feelings* of empathy and the term *prosocial norms* for social norms that showcase and prescribe prosocial *behaviour* (e.g., helping, sharing, donating). However, I recognise that empathic and prosocial norms are closely inter-related, diagnostic of one another and constitute two parts of broader normativity of prosocial orientation towards others (see also Bieberstein et al., 2021). For instance, if one sees relevant others verbally expressing empathy for outgroup members, one is also likely to infer that caring, prosocial behaviours towards outgroup members are expected, resulting in greater motivation to display both intergroup empathy and prosocial behaviour, in line with Nook and colleagues' (2016) results. Finally, the offered definition of empathic norms adheres to the classical distinction between descriptive and injunctive social norms and therefore distinguishes between *descriptive intergroup empathic norms* (i.e., how much intergroup empathy relevant others feel) and *injunctive intergroup empathic norms* (i.e., how much intergroup empathy relevant others approve or disapprove of). It is important to note that available literature has focused on both descriptive (e.g., Tarrant et al., 2009; Borinca et al., 2021a; Saguy et al., 2015) and injunctive norms (e.g., Gubler et al., 2015; Nadler & Liviatan, 2006), but no attempt has been made to compare these two types of norms. As one can imagine a discrepancy between descriptive and injunctive intergroup empathic norms (e.g., one can see an official institutional message that calls for empathy across group lines, while also witnessing unempathetic comments addressed to outgroup members online), it remains for future research to assess the independent and interactive contributions of the two norm types more systematically (see Kredentser et al., 2012; Staunton et al., 2014).

To date, multiple studies have manipulated intergroup empathic norms to increase empathy between members of groups with a history of competition, conflict, or discrimination. For instance, studies by Tarrant et al. (2009) have shown that people

empathise more with outgroup members after finding out that their ingroup also tends to show empathy for the same outgroup, thereby demonstrating *conformity to ingroup intergroup empathic norms* (see also Borinca et al., 2021a; Saguy et al., 2015). Conversely, witnessing outgroup members expressing empathy for ingroup suffering was shown to evoke empathy for the outgroup, humanisation of the entire outgroup, and willingness to reconcile, implying *reciprocity of outgroup intergroup empathic norms* (Gubler et al., 2015; Hein et al., 2016; Nadler & Liviatan, 2006). In sum, seeing intergroup empathy as normative can induce an approach motive, increasing people's willingness to empathise across group lines, which is a context in which empathy is particularly challenging.

Although the discussed findings reveal modulation of intergroup empathy depending on the availability of cues concerning ingroup or outgroup intergroup empathic norms, they do not permit strong conclusions about the subjective perceptions of these norms which arguably provide the basis for most normative effects outside the laboratory. To illustrate, imagine that an outgroup member is in need of help when no immediate cues about intergroup empathic norms are present. In this situation, empathy and prosocial behaviour are often guided by a combination of personal values and assumptions about how ingroup members would respond in equivalent situations (perceptions of ingroup intergroup empathic norms: *Other ingroup members do not really care about outgroup members, I should do the same.*) and how outgroup members would respond if the roles were reversed (perceptions of outgroup intergroup empathic norms: *They do not care about us, why should I care about them?*).

Despite the importance of understanding subjective perceptions of norms surrounding intergroup empathy and incorporating these insights when designing convincing empathy-building interventions, to my knowledge, only one study to date has directly examined these perceptions. Specifically, in post-war Sri Lanka, Penić et al. (2021) asked participants to read

vignettes describing diverse conflict experiences and to report not only the empathy they felt for the targets, but also the empathy they thought that their neighbours would feel for the same targets if they had found out about the targets' experiences. The results revealed that people tend to underestimate inclusive empathy – empathy for others facing hardship regardless of their group membership – among their ingroup members (i.e., neighbours). Although this study did not address the causal effect of perceptions of intergroup empathic norms on one's own intergroup empathy, in the context of the wider literature, it is not hard to predict that underestimating intergroup empathy among relevant others might be particularly damaging for intergroup relations. On the one hand, underestimating intergroup empathy might restrict one's own willingness to empathise across group lines (e.g., Gubler et al., 2015; Tarrant et al., 2009) and on the other hand, it might taint interpretations of actual displays of intergroup empathy, further limiting any beneficial impact of these displays on outgroup-oriented feelings and behaviours (e.g., Borinca et al., 2021b; Nadler & Liviatan, 2006; Stillerman et al., 2020).

Penić et al.'s (2021) research focused only on perceptions of ingroup intergroup empathic norms in a single intergroup environment (post-war Sri Lanka). The experiments described in the second empirical chapter of this thesis (Chapter 3) expand the focus in order to (a) provide a comprehensive description of subjective perceptions of both ingroup and outgroup empathic norms, (b) establish whether these perceptions generalise across different intergroup contexts, and (c) outline whether these perceptions vary as a function of different social factors (e.g., characteristics of empathic targets and their emotional experiences) and individual differences (e.g., individual differences in perceptions of intra- and intergroup relations) thus allowing me to move towards feature-oriented interpretations and predict when and among whom perceptions of empathic norms are most likely to be overly negative.

Building on the growing literature emphasising the primary role of norm *perceptions* over actual norms and even personal beliefs in shaping behaviour (e.g., Jachimowicz et al., 2018; Tankard & Paluck, 2016; Vallentin-Holbech et al., 2017), Chapter 3 correspondingly focuses on people's established *perceptions* of intergroup empathic norms. However, because the experiments described in Chapter 3 did not manipulate intergroup empathic norms, they cannot help to answer the question of how these norms affect one's own empathy for outgroup emotions. Thus, in Chapter 4, I return to this causal question, which has propelled most discussion on intergroup empathic norms in the literature and address it in two experiments. In the first experiment, I investigated one potential pathway through which outgroup intergroup empathic norms can stimulate empathy across group lines – through reducing how much outgroup members are perceived as responsible for and deserving of their adversities. In the second experiment, I expanded on previous research examining the influence of ingroup and outgroup intergroup empathic norms in isolation to investigate how intergroup empathy varies across social environments where cues of ingroup and outgroup intergroup empathic norms are available simultaneously and where conformity to ingroup intergroup empathic norms and reciprocity of outgroup intergroup empathic norms promote congruent or incongruent emotional reactions (i.e., apathy versus empathy). While this experiment did not manipulate features of groups and group relations as such, it was able to trace how participants' perceptions of group relations change depending on the intergroup empathic norms participants were exposed to.

Summary

In this section, I have first discussed two dominant views of empathy and explained that this thesis adheres to the narrow view, wherein empathy is defined and operationalised as vicarious experience of the empathic target's emotional state in terms of valence. I have then reviewed research from social psychology and social neuroscience which has examined how

groups influence empathy. In this review, I have pointed out that despite the sheer abundance of interdisciplinary work on the topic, there is currently limited understanding of generalised intergroup circumstances that can be expected to trigger intergroup empathy bias, regardless of the groups in question. This is because most research has focused on real-world group categories or mere ingroup/outgroup distinctions as units of explanation. An alternative approach would be to manipulate features of groups, group relations and group environments and to systematically examine different group categories along features of interest to identify when dampened empathic responses are likely to occur. Finally, I have described a plethora of interventions developed to address intergroup empathy bias and focused in particular on those which have targeted norms surrounding intergroup empathy (i.e., intergroup empathic norms) to lay the foundation for the remaining research questions addressed in Chapters 3 and 4.

Up to this point, I have reviewed research on socio-affective processes of relevance for this thesis – judgments of facial expressions of emotion (relevant for Chapter 2) and empathy (relevant for Chapters 3 and 4) – and used this research to suggest that category-oriented thinking about groups in the absence of feature-oriented thinking about groups limits broader understanding of group processes. Now, I turn to discussing the limitations of category-oriented thinking more thoroughly and to outlining how the study of group features can complement the study of group categories to yield more productive insights, applicable across different intergroup environments and across different time points.

Limitations of studying groups as group categories: How can feature-oriented thinking about groups help?

This section will start with one of the central tenets of psychological science; it depends. Specifically, whether category-oriented thinking about groups is restricting theoretical and empirical progress depends on one's research goals. Cikara (2021), one of the

proponents of feature-oriented thinking about groups[14], suggested that if one's goal is to understand a *specific intergroup context in the here and now* (e.g., Israeli-Palestinian conflict), one could focus narrowly on the group categories in question, discussing the particularities of their relations and histories to provide in-depth, but non-generalisable predictions. However, if one's goal is to understand *groups and group relations more broadly* and to make generalisable predictions about novel intergroup contexts that might not even exist at present[15] and about existing intergroup contexts in the future (e.g., Israeli-Palestinian relations in a hundred years), one would benefit more from trying to identify abstract features of groups, group relations and group environments that are predictive of different outcomes, regardless of the specific groups in question (Cikara, 2021; also see Cikara et al., 2022; Cikara & Van Bavel, 2014). Since the latter goal better captures the goal of this thesis, using existing group categories (e.g., "Croatian" versus "Serbian") or abstract group categories ("ingroup" versus "outgroup") as units of explanation without discussing broader contextual features that might be of relevance to the observed effects is unlikely to get me far in the pursuit of understanding *what kind of groups and intergroup environments* are likely to influence judgments of facial expressions or empathy.

Cikara (2021) lists multiple limitations of category-oriented thinking about groups. First, group categories are not set in stone; multiple group categories can evolve into a bigger

[14] Again, the terms *category-oriented thinking* and *feature-oriented thinking about groups* were introduced in this thesis.

[15] Examples of recent social identities (i.e., group categories) that were not relevant a few decades or even years back are (a) the "South Sudanese" identity, which became salient after the country's formation in 2011 and is still in the process of development (Frahm, 2012), (b) the "Kosovar" identity, which was triggered by the 1998-1999 Kosovo war (Maloku et al., 2016), and (c) "Leaver" and "Remainer" identities, which only became influential in the UK around the 2016 referendum (Hobolt et al., 2021). Similarly, social identities that were very salient a few decades back – "West German", "East German", or "Yugoslav" – are now not as influential as they used to be (even though they might still resonate with some people).

group category over time (e.g., unification of individual German-speaking states to form a German Empire which ultimately removed identification with individual states to leave only the overarching "German" identity) or a bigger group category can be expanded to absorb new group categories (e.g., Italian and Irish immigrants becoming "White" in the early 20th century, Cikara, 2021). Similarly, a bigger group category can break down into multiple smaller ones (e.g., "Yugoslav" identity broke down into individual "Croatian", "Bosnian", "Macedonian", "Montenegrin", "Serbian", and "Slovenian" identities).

The former example of smaller group categories evolving into a bigger group category might seem close to the concept of *recategorisation*, wherein group members' perceptions of "us" and "them" are replaced with more inclusive "we" (Gaertner et al., 2000). This could mean, for example, that members of different ethnic groups embrace their shared national identity over their respective ethnic identities. However, while recategorisation is only concerned with changing perceivers' focus in an environment where multiple group categories can be active, the examples I started with refer to actual instances of new group categories surfacing or old group categories fracturing. This surfacing and fracturing cannot be easily understood if one maintains the explanatory focus on group categories (i.e., social identities) instead of contextual factors and relational dynamics contributing to the changing group category landscape. Second, even when group categories remain largely unchanged, the relations between them are prone to change (e.g., just compare relations between European countries during the two World Wars to their relations in 2022). Category-oriented interpretations rarely account for intergroup relations, rendering these interpretations unable to explain changes in intergroup feelings and behaviours that follow changes in relational dynamics.

While the first two limitations mostly apply to real-world groups, the final limitation I am about to summarise applies to minimal groups and was already discussed in more detail

earlier in this chapter. Specifically, minimal intergroup biases are most frequently interpreted as a product of *mere* group category membership and its positive valuation (Dunham, 2018), yet this membership is often experimentally confounded with other potentially relevant features (e.g., similarity, coalitional relations, common fate, Yamagishi & Kiyonari, 2000) and even when it is not, it can be construed as a black-box container of people's assumptions about what it means to be "us" (Schiffman & Wicklund, 1992)[16]. In other words, the surface-level meaninglessness of minimal groups should not lead one to believe that all that is left to them is *mere* group category membership. Instead, it is more likely that all that is left to minimal groups is the meaning people attribute to specific group category labels (e.g., "overestimators" versus "underestimators", Hong & Ratner, 2020) and to group category sharing ("ingroup members" versus "outgroup members", Misch et al., 2021). Thus, instead of treating minimal group category membership as a self-contained explanation of any observed effects, it would be more productive to unpack the underlying features of minimal groups, and examine which of these features, once removed, eliminates intergroup bias.

In summary, category-oriented thinking about groups puts group categories (e.g., racial groups) ahead of features (e.g., power relations, facial features) thus making it difficult to generalise any findings from one intergroup context to another and systematise the conditions under which intergroup biases are likely to occur. The dominance of category-oriented thinking in the fields of social cognition and social neuroscience was partly propelled by the minimal group paradigm and associated interpretations offered by the Social Identity Perspective, which somewhat "equated social identity with the physical presence in

[16] One anecdotal illustration of how people imbue minimal groups, particularly minimal *ingroups*, with rich meaning involves a participant who, after being assigned to a minimal group and finding out that sharing the same minimal group with someone does not necessarily signify cooperative relations (Experiment 3 in Chapter 2), said: "What do you mean I will be in the same group with someone *and* compete with them? How is that even possible? The groups do not make any sense then."

the group [category]" (Schiffman & Wicklund, 1992, p. 29), as "though tagging a person with a [group] category comprehensively captures their thoughts, feelings and perceptions, and how others will treat them and why" (Cikara et al., 2022, p. 537). However, as mentioned earlier in this chapter, it is important to note that an abundance of social psychological research inspired by the Social Identity Perspective explicitly examined how different features of groups, group relations and broader social environments in which groups operate affect the occurrence and the size of intergroup biases (Giannakakis & Fritsche, 2011; Jackson et al., 1996; Mullen et al., 1992). Therefore, the discussed limitations of category-oriented thinking about groups mostly apply to social cognition and social neuroscience research which tended to equate "groups" with "group categories" and examine how these "group categories" (e.g., race, gender, minimal groups), in isolation from contextual features, affect psychological processes and behaviours (e.g., emotion judgments, empathy, prosocial behaviours). Now, I will provide examples of how combining the study of group categories – both minimal and real-world – with the study of group features can address these limitations and yield more productive insights.

Examples of feature-oriented thinking in social psychological research

In this section, I will offer two examples of feature-oriented thinking in action. The first example will illustrate research which has embraced feature-oriented thinking in the domain of intergroup empathy by directly manipulating features of groups and group relations (Cikara et al., 2014). The second example will illustrate research which has embraced feature-oriented thinking in the domain of intergroup violence (i.e., hate crimes, Cikara et al., 2022) by examining multiple real-world group categories in a way that focuses on these categories' shared features as units of explanation.

First, Cikara and colleagues (2014) sought to identify general conditions under which people fail to empathise with outgroup members. Specifically, they manipulated the

following features: (a) coalitional relations between groups (i.e., whether the two minimal groups cooperated, competed or were independent from each other), (b) competitive threat (i.e., whether the minimal outgroup posed a threat to minimal ingroup's superiority) and (c) perceived entitativity (i.e., whether the two minimal groups were perceived as completely separate or greatly overlapping entities). The results revealed that intergroup empathy bias was greatest when groups were asked to compete and smallest or completely absent when groups were asked to cooperate. Furthermore, once competitive relations were instantiated, whether the minimal outgroup was still a threat to minimal ingroup's victory or the minimal ingroup had already won the competitive game did not change the size of participants' intergroup empathy bias. This illustrated that an acute sense of competitive threat mattered less for intergroup empathy than general competitive relations. Finally, intergroup empathy bias was larger when minimal groups were portrayed as separate as opposed to overlapping entities. This research perfectly captures the value of feature-oriented thinking; instead of focusing on "minimal ingroups" and "minimal outgroups" as such, Cikara and colleagues (2014) provided insights into the abstract features of groups (e.g., their entitativity) and group relations (e.g., their coalitional implications) that can amplify or attenuate intergroup empathy bias, allowing future research to build on these insights more easily across a variety of intergroup contexts.

While the first example involved direct manipulation of group features, the second example illustrates how features can be put to the forefront even when working with real-world group categories. Specifically, Cikara and colleagues (2022) used census data to identify factors which predict the prevalence of hate crimes directed towards different minority groups across the UK and the US (e.g., Arab/Muslim, Asian, Black, and Hispanic people). The results revealed that as a particular group increased in rank size relative to other groups, its members became a more frequent target of hate crimes, a finding which persisted

even after accounting for a plethora of confounds and alternative explanations (e.g., baseline group size). Even though Cikara and colleagues (2022) analysed data which included real-world group categories with their unique stereotypical features and histories, rank size of a particular group in comparison to other available groups was shown to represent a *group-invariant feature* predictive of hate crimes. The group-invariant nature of this finding makes it easier for researchers to build on it in any part of the world (e.g., any of the 41 states which submitted hate crime data to the Office for Democratic Institutions and Human Rights, ODIHR, 2021) and at any time in the future (e.g., ten or a hundred years from now).

Summary

In this section, I joined a growing number of researchers who have advocated moving beyond group categories as units of explanation and incorporating contextual features into one's research and theorising (Cikara, 2021; Cikara et al., 2022; Cikara & Van Bavel, 2014; Sen & Wasov, 2016; see also Brubaker, 2006; Gillespie et al., 2012). I started by discussing how category-oriented thinking about groups is limited in its ability to produce insights that are generalisable across space and time. I then emphasised how, in the context of the minimal group paradigm, category-oriented thinking about groups obscures the fact that minimal groups are not empty group categories, but more likely containers of people's intuitions about what it means to be "us". I continued to outline how feature-oriented thinking about groups prioritises group features over group categories and how the study of group features can operate in the context of known group categories. Specifically, I reviewed two examples of research championing feature-oriented thinking about groups by either (a) manipulating features of minimal groups and their relations to identify *what kind of groups and intergroup environments* are most likely to trigger intergroup bias (Cikara et al., 2014) or by (b) examining multiple real-world groups to identify generalised group features predictive of group members' treatment (i.e., the prevalence of hate crimes, Cikara et al., 2022). In the

next and final section, I will provide a brief overview of the current thesis and outline how the included experiments model feature-oriented thinking about groups.

Overview of the current thesis

In this chapter, I have reviewed relevant literature, introduced the analytic lens adopted in this thesis (feature-oriented thinking about groups) and listed research questions addressed across seven experiments included in three empirical chapters (Chapters 2-4). Even though most experiments in this thesis investigate the influence of group relations alongside group category membership on judgments of facial expressions and empathy, each chapter does not always pick up on the findings from the preceding one. Nevertheless, what unites all experiments is their effort to move beyond group category membership and offer more nuanced feature-oriented interpretations, thus allowing predictions regarding *when* intergroup biases in judgments of facial expressions and empathy are likely to occur, regardless of the groups under consideration.

Chapter 2 reports three experiments (Experiments 1-3) which build on the work of Lazerus and colleagues (2016) who found that people tend to judge facial expressions of (cooperative) minimal ingroup members as conveying more positive feelings than expressions of (competitive) minimal outgroup members. To examine whether this result can be better understood as a product of group [category] membership, co-occurring coalitional relations (i.e., cooperation, competition) or their interaction, I orthogonally manipulated group membership of, and coalitional relations with, the expresser to examine how these two factors independently and jointly influence valence, arousal, and genuineness judgments of facial expressions (Experiments 1-2). Furthermore, I examined whether group membership of, and coalitional relations with, the expresser lead to differential visual processing of facial expressions, and whether this differential visual processing might underlie the influence of these two factors on judgments of facial expressions (Experiment 3). In sum, Experiments 1-

3 directly examine whether previously documented minimal ingroup biases in valence judgments of facial expressions can be re-cast as a product of co-occurring coalitional relations.

Chapter 3 includes three experiments (Experiments 4-6a) which seek to understand people's unaltered perceptions of empathic norms in intergroup contexts, the accuracy of these perceptions, and their dependency on contextual information about empathic targets and their emotional experiences as well as on individual differences in how sampled group relations are perceived. Specifically, across ethnic, political and minimal intergroup contexts, Chapter 3 addresses the following research questions: (a) how much empathy are ingroup and outgroup members assumed to feel towards own-group versus other-group empathic targets (Experiments 4-6a); (b) how the levels of intergroup empathy bias participants assume that ingroup and outgroup members show compare to the levels of intergroup empathy bias participants report themselves (Experiments 4-6a) and (c) to the levels of intergroup empathy bias ingroup and outgroup members in fact report on average (Experiments 4-5); (d) whether participants assume that average ingroup and outgroup members discriminate equally in their empathy (Experiments 4-6a); (e) whether the availability of different contextual information influences how much intergroup empathy bias ingroup and outgroup members are assumed to show (Experiments 4-5); and (f) whether participants' general perceptions of group relations predict their perceptions of empathic norms (Experiments 4-5). In sum, Experiments 4-6a both manipulated features of the context (i.e., features of empathic targets and their emotional experiences) and assessed people's perceptions of group relations to understand when and among whom perceptions of intergroup empathic norms are most likely to be negative and inaccurate.

Chapter 4 includes two experiments (Experiments 6b-7) which examine the causal influence of intergroup empathic norms on one's own intergroup empathy. Experiment 6b

examined whether outgroup intergroup empathic norms encourage intergroup empathy by reducing how much outgroup members are subsequently perceived as deserving of their adversities. Experiment 7 examined how people navigate intergroup contexts where both ingroup and outgroup empathic norms are simultaneously available and are either aligned (i.e., both promote intergroup apathy or empathy) or misaligned (i.e., one promotes intergroup apathy and the other one intergroup empathy). Given that no previous research has investigated the mechanisms through which intergroup empathic norms exert their influence or the preferences people display when faced with two irreconcilable intergroup empathic norms, Experiments 6b and 7 did not manipulate any group features on top of manipulating intergroup empathic norms. Instead, Experiments 6b and 7 aim to be transparent about the group features implied by the experimental design and to discuss potentially relevant group features whose influence should be addressed in future research. Finally, this thesis concludes with a general discussion, covered in Chapter 5, in which I summarise the main findings, outline future directions as well as strengths and limitations of present work.

CHAPTER 2: THE INFLUENCE OF GROUP CATEGORY MEMBERSHIP AND COALITIONAL RELATIONS ON JUDGMENTS OF FACIAL EXPRESSIONS OF EMOTION

Abstract

Real-world groups have long been known to influence categorical recognition and continuous judgments of facial expressions of emotion, with more recent work suggesting that even minimal groups might lead to similar biases. However, in past experiments employing a minimal group paradigm, coalitional relations typically involved explicit or assumed intragroup cooperation alongside intergroup competition, making it difficult to parse whether intergroup biases in judgments of facial expressions can be explained as a product of group category membership or co-occurring coalitional relations. Thus, across three experiments, I investigated how minimal group membership and coalitional relations influence valence, arousal, and genuineness judgments of facial expressions of emotion, and whether there are any identifiable attentional and physiological markers associated with processing of social information in intergroup contexts. In Experiments 1 and 2, across angry, ambiguous, and happy facial expressions, I observed an *ingroup positivity bias*, wherein ingroup members' facial expressions were judged as conveying more positive emotions than outgroup members' facial expressions, and a *cooperation positivity bias*, wherein facial expressions of people linked to a cooperative context were judged as conveying more positive emotions than facial expressions of people linked to a competitive context. Experiment 3 employed eye-tracking and found no evidence that group category membership and coalitional relations affect valence judgments of facial expressions by altering how people allocate their visual attention to facial features. This result makes it more probable that ingroup and cooperation positivity biases arise post-perceptually, as a consequence of

affective priming, wherein the evaluative response to ingroup members and cooperation spills over to the domain of valence judgments of facial expressions. Finally, Experiment 3 revealed patterns of pupil dilation which dissociated unexpected group-coalition pairings (i.e., ingroup-competition, outgroup-cooperation) from expected ones (i.e., ingroup-cooperation, outgroup-competition) and patterns of skin conductance response which indexed sensitivity to anticipated competition. Taken together, these results suggest that both group category membership and coalitional relations independently influence valence judgments of facial expressions, but that their influence is unlikely to operate through triggering top-down attentional shifts.

Introduction

In Chapter 1, I reviewed social psychological research which has investigated how groups influence categorical recognition and continuous judgments of facial expressions of emotion. To recap, this literature has documented ingroup-favouring biases related to accuracy (e.g., Elfenbein & Ambady, 2006), confidence (e.g., Beaupré & Hess, 2006), intensity (e.g., Kommattam et al., 2019), and valence (e.g., Kunstman et al., 2016) of judgments of facial expressions of emotion in real intergroup contexts, with more recent work suggesting that some of these biases occur even in minimal intergroup contexts (e.g., Lazerus et al., 2016; Young & Wilson, 2018). In one such minimal group experiment, Lazerus and colleagues (2016) found that people tend to interpret facial expressions of minimal ingroup members as conveying more positive emotions than equivalent expressions of minimal outgroup members[17]. As discussed in Chapter 1, since Lazerus and colleagues (2016) set minimal groups in direct competition, it remains an open question whether competition (i.e., negative interdependence) is a necessary condition for *ingroup positivity bias* in judgments of facial expressions. If this is the case, introducing cooperation (i.e., positive interdependence) between minimal groups should eliminate the observed bias (see also Cikara et al., 2014). In a series of experiments presented in this chapter, I set out to address this open question. Specifically, I directly manipulated *coalitional relations* with minimal outgroup members to examine how they affect ingroup positivity bias in judgments of facial expressions.

[17] In one exploratory analysis, Lazerus et al. (2016) also showed that minimal ingroup members' facial expressions of emotion are interpreted as conveying higher levels of arousal than minimal outgroup members' facial expressions of emotion. Given that this finding was reported only once and was not of primary interest in Lazerus and colleagues' (2016) work, valence judgments of facial expressions will be the focus of this introduction and subsequent hypotheses.

Another easily overlooked methodological decision of Lazerus and colleagues (2016) warrants further consideration, as it connects to the broader discussion of how ingroup biases in minimal group experiments should be interpreted. Specifically, in their introduction of intergroup competition, the authors also introduced intragroup cooperation by making participants believe that they would cooperate with other minimal ingroup members against minimal outgroup members to secure bonus payment. While such explicit "group-coalition" confounding is rarely seen in the literature (but also see Bruneau et al., 2015)[18], most minimal group experiments fail to adequately discuss that group category membership and coalitional relations might be implicitly confounded. For instance, it has been shown that people tend to hold a generalised expectation of within-group cooperation (Balliet et al., 2014; Brewer, 2008; Yamagishi & Kiyonari, 2000). This "natural" confounding would not be a problem had the ingroup biases in minimal group experiments been interpreted in a way that openly acknowledged that minimal groups might in fact track explicit or assumed coalitional allegiances. Over the years, however, ingroup biases in minimal group experiments have predominantly been interpreted in terms of self-categorisation with the ingroup (Otten, 2016) with only a few authors emphasising that the minimal group paradigm often does not remove interdependence between ingroup members, rendering coalitional relations a candidate underlying feature that gives minimal groups their motivational potency (Gaertner & Insko, 2000; Rabbie et al., 1989; Yamagishi & Kiyonari, 2000).

[18] Due to reliance on bogus personality tests to create minimal groups (see Young & Hugenberg, 2010 for an example), past research more often introduced confounding between group category membership and intrinsic similarity, implying that minimal ingroup members are intrinsically similar to each other and dissimilar from minimal outgroup members. Within-group intrinsic similarity is not a necessary condition for defining and recognising group categories, but might, however, imply cooperative within-group relations (see Fischer, 2009 and Toma et al., 2012 for a discussion of how perceived similarity might be related to cooperation).

Recent work from Misch and colleagues (2021) has found that coalitional relations can at least partly explain ingroup biases considered to be the bedrock of the minimal group paradigm. Specifically, the authors showed that people's tendencies to report greater liking of minimal ingroup members than minimal outgroup members and to preferentially allocate resources to minimal ingroup members relative to minimal outgroup members disappear when minimal group membership and anticipation of cooperation are decoupled. In other words, once cooperation was anticipated with outgroup members as opposed to ingroup members, all ingroup biases people initially expressed in preference and resource allocation measures were eliminated.

Going back to the work of Lazerus and colleagues (2016), due to their experiment set-up, ingroup positivity bias in judgments of facial expressions of emotion can be interpreted through the lens of coalitional relations rather than group category membership. In particular, participants could have simply judged facial expressions of those with whom they were to compete (who also happened to be labelled as minimal outgroup members) as conveying more negative emotions than facial expressions of those with whom they were to cooperate (who also happened to be labelled as minimal ingroup members). If supported, this alternative explanation would mean that the observed *ingroup positivity bias* would be better named *cooperation positivity bias*. Thus, to address the question of whether biases in valence judgments of facial expressions are in fact driven by coalitional relations or whatever remains of group category membership once coalitional relations are taken out, experiments in Chapter 2 independently manipulated the two factors by introducing not only *intergroup* competition and cooperation, but also *intragroup* competition and cooperation.

It is worth noting that explicit judgments of another person's facial expressions represent only the end-result of a dynamic interaction between bottom-up cues (e.g., facial features) and top-down influences (e.g., prior knowledge about the expresser's group

membership and coalitional relations with them; Brooks & Freeman, 2018; Freeman & Ambady, 2011; Freeman & Johnson, 2016). As such, these judgments alone may not always provide sufficient insight into the processing mechanisms that preceded them (see Freeman, 2018) and might even conceal influences that may be apparent in more implicit measures. For this reason, a number of previous studies has combined self-report judgments with real-time behavioural, physiological, and/or neural measures (e.g., mouse-tracking, pupillometry, EEG) to investigate the influence of group processes on social perception. For instance, Stolier and Freeman (2016) showed that stereotypical associations between group and emotion categories (e.g., men-anger) shape social perception by looking at how participants' mouse trajectories changed during a social categorisation task and by examining whether such stereotypical associations altered the neural patterns in the fusiform and orbitofrontal cortices, regions previously shown to integrate perceptual representations with top-down expectations (Bar et al., 2006; Freeman et al., 2015). In Stolier and Freeman's (2016) work, even though participants' explicit responses in the social categorisation task were not informative of the underlying mechanisms, participants' mouse trajectories were able to reveal the influence of stereotypical group-emotion associations. Similarly, a different experiment sought to investigate whether certain group-emotion pairings are recognised as incongruent (e.g., ingroup-anger) and whether such conceptual incongruence impacts processing of ingroup and outgroup members' facial expressions (Gamond et al., 2017). Whereas group-emotion incongruence made no difference in how accurate or fast participants' recognition of facial expressions was, as indexed by self-reported judgments and reaction time respectively, group-emotion incongruence was indeed shown to affect early visual processing of facial expressions of emotion. Specifically, expressions incongruent with prior expectations (i.e., ingroup anger and outgroup joy) elicited larger EEG activity 30 to 100 ms after onset in comparison to expressions congruent with prior expectations (i.e.,

ingroup joy and outgroup anger), which was interpreted as a sign of increased attention to unexpected stimuli (Krall et al., 2015). Importantly, in both of these examples, adoption of unobtrusive and real-time measures resulted in important insights which could not be gleaned from participants' judgments of facial expressions alone (see Amodio et al., 2003; Cikara & Fiske, 2012; Xu et al., 2009 for examples of similar decoupling between self-report judgments and physiological or neural measures).

In current work, I sought to get a more mechanistic understanding of how social information about the expresser – the expresser's group category membership and coalitional relations with them – may drive the processing of their facial expressions. Specifically, I examined whether visual attention, as indexed by eye-tracking, and physiological arousal, as indexed by pupil dilation and skin conductance response, might reveal covert influence of group category membership and coalitional relations which may or may not correspond to the influence these factors have on overt judgments of facial expressions. In particular, the adoption of these measures allowed me to test two particularly important possibilities. First, I used eye-tracking to examine whether group category membership and/or coalitional relations affect judgments of facial expressions by altering how people allocate their visual attention to salient facial features. Addressing this question was particularly important because it has been hypothesised that factors such as group category membership and coalitional relations only come into play "post-perceptually", affecting judgments of otherwise similarly explored facial expressions (see Lazerus et al., 2016). Second, as changes in physiological arousal are known to index people's responsiveness to unexpected information (Cuve et al., 2018; Dindo & Fowles, 2008; Hepach & Westermann, 2013; Mathersul et al., 2013), pupil dilation and skin conductance response were used to index implicit processing of social information about expressers (see also Brooks & Freeman, 2018; Gamond et al, 2017). Specifically, given that features such as cooperation normatively co-

occur with sharing a group category membership, I assessed whether people show greater physiological responsivity to incongruent pairings of social information about the expresser (e.g., ingroup-competition), regardless of whether this incongruence is evident in participants' judgments of facial expressions. The use of eye-tracking and peripheral physiology was also motivated by the practical potential of these measures, given their greater accessibility and ease of set-up for long behavioural experiments as compared to neural measures. Finally, specific hypotheses related to visual attention to faces and physiological arousal will be discussed in more detail in the introduction of the relevant experiment (Experiment 3).

This chapter is divided into three sections, each corresponding to a single experiment. In Experiment 1, I examined how group category membership of and coalitional relations with the expresser influence valence, arousal, and genuineness judgments of their angry, ambiguous, and happy facial expressions. In Experiment 2, I removed minor methodological limitations of Experiment 1 and addressed the same research question using a larger set of ambiguous facial expressions. Finally, in Experiment 3, I examined whether social information about the expresser leads to differential visual processing of facial expressions, and whether any indices of this differential visual processing might underlie the influence of group category membership and/or coalitional relations on judgments of facial expressions. Simultaneously, Experiment 3 adopted physiological measures to assess whether perceivers are sensitive to incongruent social information about the expresser. Within each section, I first detail the methods and the results of the experiment, followed by a summary and a discussion of experiment-specific findings. The chapter concludes with a general discussion, in which I connect the findings from the three experiments to prior research, outline future directions as well as strengths and limitations of present work.

Hypotheses

In this section, I will briefly outline hypotheses for Experiment 1 which pertain to how valence judgments of facial expressions might be affected by group category membership of the expresser, coalitional relations with the expresser or the interaction between these two factors.

H1: Participants will judge facial expressions of ingroup members as conveying more positive emotions than facial expressions of outgroup members, regardless of the intra- or intergroup coalitional relations in place (*ingroup positivity bias*).

H2: Participants will judge facial expressions of those with whom they are to cooperate as conveying more positive emotions than facial expressions of those with whom they are to compete, regardless of their ingroup or outgroup status (*cooperation positivity bias*).

H3: Participants will judge facial expressions of ingroup members with whom they are to compete as conveying the most negative emotions, as these members violate the ingroup cooperation norm. However, it is important to note that competition and cooperation in current experiments were not a reflection of participants' choices or supposed ingroup and outgroup members' choices, but rather a consequence of random allocation (e.g., a specific ingroup member was randomly assigned to interact competitively with the participant, they did not choose to be competitive). This feature makes the third hypothesis less likely as ingroup competition might not be perceived as a norm violation in the absence of perceived intentionality (Ma et al., 2016; Wang et al., 2016).

H4: All listed hypotheses (H1-H3) were addressed in the context of both prototypical and ambiguous facial expressions. Whichever of the three effects (if any) is observed is expected to be more pronounced when facial expressions of emotion are ambiguous as

opposed to obvious (i.e., prototypical facial expressions of anger and happiness) because the lack of clarity and precision in the sensory input should leave more scope for the influence of group category membership and/or coalitional relations (Freeman et al., 2020).

Experiment 1

Method

Participants. Eligible participants had to reside in the UK, be at least 18 years old, have no formal education in psychology, have no history of diagnosed mental health disorders or autism and have normal colour vision. Participants were recruited using the university participant network (N = 36) as well as Prolific (N = 134). Two Prolific participants were excluded because they provided inaccurate attention check responses on more than 20% of the trials in the emotion task, resulting in the final sample of 168 adults (46.4% women, 53.6% men) between the ages of 18 and 60 (M_{AGE} = 24.90, SD_{AGE} = 7.08),

Experiment 1 was substantially impacted by the Covid-19 pandemic. Specifically, participant recruitment for Experiment 1 started in late January 2020 and this experiment was originally envisioned to take place in person and include eye-tracking and physiological measures alongside explicit judgments of facial expressions of emotion. By March 2020, when the University of Oxford announced its full closure, a total of 9 participants had been recruited. In May 2020, after it had become clear that in-person testing would not be resuming soon, Experiment 1 was moved online where 134 more participants were recruited on Prolific (www.prolific.co). Given the online nature of the study, the 134 participants provided only their explicit judgments of facial expressions of emotion with no eye-tracking and physiological data collected. Finally, since the original in-person sample size was insufficient to yield any valid conclusions in the domain of eye-tracking and peripheral physiology, another 27 participants completed the lab-based version of the experiment in

September 2022. The experiment set-up remained (almost) identical across online and lab-based testing[19], allowing for all participants' valence, arousal, and genuineness judgments of facial expressions of emotion to be combined into a single dataset and analysed jointly, resulting in a full sample of 168 participants after exclusions. On the other hand, eye-tracking and physiological data of lab-based participants recruited in 2020 (N = 9) and 2022 (N = 27) were also combined and are reported in this thesis under Experiment 3.

Power analysis. Because no prior work directly investigated the interplay between group category membership and coalitional relations in the domain of emotion judgments, Experiment 1 did not test only one hypothesis and a precise a priori power analysis focusing on a specific effect of interest was thus not possible. However, to get an approximate estimate of the required sample size, I conducted simulation-based power analyses (1000 iterations) for factorial ANOVA designs in the R package Superpower (Lakens & Caldwell, 2021). In the simulated data, either group category membership or coalitional relations had a main effect on participants' valence judgments of facial expressions and this effect interacted with emotion category such that participants displayed greater bias in their valence judgments when facial expressions were ambiguous as opposed to prototypical (see Table 2.1 for assumed mean and standard deviation values). This analysis revealed that a sample of 120 participants should suffice to identify the simulated main effect with maximal (100%) statistical power and the simulated interaction effect with 82% statistical power at alpha = .05. In addition to this a priori power analysis, I conducted post hoc sensitivity simulations for linear mixed models (1000 iterations, Green & Macleod, 2016) which revealed that our

[19] The differences between lab-based and online versions of Experiment 1 mostly concerned the order in which different measures were distributed as well as the duration of inter-trial intervals and the intervals between different segments of the same trial (i.e., inter-segment intervals). These differences are described in more detail in a latter section.

observed group- and coalition-related estimates were 24% and 6.5% larger respectively than the minimal detectable estimate with 80% power.

Table 2.1

Means and standard deviations used in the simulation-based power analysis

Emotion	Group	Coalitional relations	Assumed mean of valence judgments	Assumed standard deviation of valence judgments
Angry	Ingroup	Competition	30	5.5
Angry	Ingroup	Cooperation	30	5.5
Angry	Outgroup	Competition	29.5	5.5
Angry	Outgroup	Cooperation	29.5	5.5
Ambiguous	Ingroup	Competition	51.5	5.5
Ambiguous	Ingroup	Cooperation	51.5	5.5
Ambiguous	Outgroup	Competition	50	5.5
Ambiguous	Outgroup	Cooperation	50	5.5
Happy	Ingroup	Competition	70	5.5
Happy	Ingroup	Cooperation	70	5.5

Happy	Outgroup	Competition	69.5	5.5
Happy	Outgroup	Cooperation	69.5	5.5

Notes. Assumed means and standard deviations in each experimental condition based on which simulation-based power analysis was conducted. Here, data were simulated such that group category membership has a main effect on participants' valence judgments of facial expressions. This main effect is qualified by a two-way group category membership x emotion category interaction which reflects the expectation that any bias participants show will be greater when facial expressions are ambiguous as opposed to prototypical.

Facial stimuli. Angry and happy facial expressions from the Chicago Face Database (CFD, Ma et al., 2015) and the Radboud Faces Database (RFD, Langner et al., 2010), which showed sufficient activation of relevant action units (see "Facial stimuli selection" below, Ekman & Friesen, 1978; Ekman et al., 1980; Gosselin et al., 2010) as detected by OpenFace (Baltrusaitis et al., 2018), were morphed in Psychomorph (Sutherland et al., 2017) to create ambiguous expressions. Having chosen the facial stimuli, all images were equalised for luminance and spatial frequency using the SHINE toolbox in MATLAB (Willenbockel et al., 2010) and non-facial areas were cropped (see Figure 2.1).

The exact number of facial stimuli varied depending on where and when participants completed the experiment. Specifically, initial lab-based participants (N = 9) rated angry, ambiguous, and happy facial expressions of 48 targets (24 female, 24 male), which corresponds to a total of 144 facial stimuli. Online participants and participants who completed the lab-based version of the experiment in 2022 (N = 161) rated angry, ambiguous, and happy facial expressions of 32 targets (16 female, 16 male), which corresponds to a total of 96 facial stimuli. Regardless of the exact number of facial stimuli, each experimental condition included an equal number of male and female targets, and target identities were randomly assigned to an experimental condition for each participant (e.g., participant 1 could see a specific female target in the "ingroup competition" condition while participant 2 could see the same target in the "outgroup cooperation" condition).

Facial stimuli selection. Facial stimuli used in Experiment 1 were selected in three steps. In the first step, I selected expressers from the CFD (Ma et al., 2015) and the RFD (Langner et al., 2010) based on the following criteria: (a) all expressers needed to be White adults, (b) all expressers needed to be frontally photographed, (c) all expressers needed to provide facial expressions of both anger and happiness and the facial expression of happiness needed to include a smile with visible teeth. A total of 109 expressers (CFD: 70 expressers, RFD: 39 expressers) met all criteria and proceeded to the next selection step.

While the authors of RFD (Langner et al., 2010) explicitly relied on the classic Facial Action Coding System (Ekman et al., 2002) to generate facial expressions, this was not the case with facial expressions taken from CFD (Ma et al., 2015). Thus, in the second step, I ran all facial expressions through OpenFace (Baltrusaitis et al., 2018), an open-access software which detects action units (0 = action unit absent, 1 = action unit present) and estimates their intensity on a five-point scale. Expressers were included if OpenFace estimates met the following criteria: (a) action units 4, 5, 7, 17, 23, and 24 had to be labelled as present in the facial expression of anger, (b) the intensity of action unit 4 (i.e., brow lowerer) had to be above 1.5, while the intensity of action units 6 (i.e., cheek raiser) and 12 (i.e., lip corner puller) had to be below 1 in the facial expression of anger, (c) action units 6, 12, and 25 had to be labelled as present in the facial expression of happiness, and (d) the intensity of action unit 4 (i.e., brow lowerer) had to be below 1, while the intensity of action units 6 (i.e., cheek raiser) and 12 (i.e., lip corner puller) had to above 1.5 in the facial expression of happiness. The specific action units were chosen to match action units targeted during the development of RFD (Langner et al., 2010). The specific cut-off values (i.e., 1 and 1.5) for intensity estimates were based on the distribution of intensity estimates across all facial expressions of anger and happiness respectively and were set to exclude facial expressions of anger and happiness which showed no or very few signs of prototypically expected action units. Setting

these cut-off values allowed me to minimise within-emotion-category differences and maximise prototypical between-emotion-category differences[20]. Additionally, creating a similar pattern of action unit intensities across facial expressions within the same emotion category was particularly important given that these facial expressions were subsequently used to create ambiguous facial expressions. Ultimately, this second step led to an exclusion of 22 expressers, resulting in a final sample of 87 expressers who provided facial expressions of both anger and happiness.

In the third step, I created a 9-step morphing sequence between the prototypical facial expression of anger and the prototypical facial expression of happiness using Psychomorph (Sutherland et al., 2017). From each morphing sequence, I then selected the middle three morphs and validated them in a pilot study. In this pilot study (N = 45, 82.22% female, 17.78% male, M_{AGE} = 25.09, SD_{AGE} = 8.79), participants came to the lab to complete a single task where they rated ambiguous facial expressions' valence, arousal, and genuineness on 100-point scales. Given that 261 ambiguous facial expressions (87 expressers x 3 morph levels) would take two hours to rate, participants were instead assigned to rate one of the four subsets of ambiguous facial expressions (N_{SET1} = 13 participants, N_{SET2} = 10 participants, N_{SET3} = 11 participants, N_{SET4} = 11 participants). Once the pilot study had been completed, descriptive statistics were calculated for each included facial expression. The final set of ambiguous facial expressions was selected based on median valence and arousal which were ideally intended to cover the mid-range of 100-point rating scales and based on median genuineness which was intended to be as high as possible. A total of 48 ambiguous facial expressions, alongside their corresponding angry and happy facial expressions, were selected

[20] This is not to imply that I adopt the view that there is a one-on-one mapping between an emotional experience (e.g., experience of anger) and a particular facial expression. However, since the paradigm adopted in Experiments 1-3 was very subtle and highly likely to result in small effects, if any, facial expressions had to be carefully pre-selected to better map and control for their bottom-up variability.

for lab-based testing in 2020[21], while this list was subsequently shortened to 32 to accommodate the switch to online testing[22].

Figure 2.1

Example stimuli used in Experiment 1



Procedure. Since lab-based and online versions of Experiment 1 differed in a few places, their corresponding procedures will be described separately. Their differences will then be explicitly highlighted to avoid any confusion.

Online procedure. The online experiment was hosted in Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), and was framed as if it included two parts, with participants completing the first part of the experiment and expecting the second part to take place within the following week. In the (supposed) first part of the experiment, participants were randomly

[21] Validation estimates for the 48 ambiguous facial expressions: $M_{VALENCE}$ = 41.17-61.09; $Median_{VALENCE}$ = 41.00-59.00; $SD_{VALENCE}$ = 7.02-25.42; $M_{AROUSAL}$ = 29.27-64.82; $Median_{AROUSAL}$ = 30.00-66.00; $SD_{AROUSAL}$ = 10.77-26.48; $M_{GENUINENESS}$ = 29.18-77.75; $Median_{GENUINENESS}$ = 30.00-82.50; $SD_{GENUINENESS}$ = 15.22-32.54

[22] Validation estimates for the 32 ambiguous facial expressions: $M_{VALENCE}$ = 42.10-59.91; $Median_{VALENCE}$ = 43-55.50; $SD_{VALENCE}$ = 7.02-19.36; $M_{AROUSAL}$ = 29.27-64.82; $Median_{AROUSAL}$ = 30.00-66.00; $SD_{AROUSAL}$ = 11.03-24.30; $M_{GENUINENESS}$ = 29.18-70.33; $Median_{GENUINENESS}$ = 30.00-76.00; $SD_{GENUINENESS}$ = 15.22-32.54

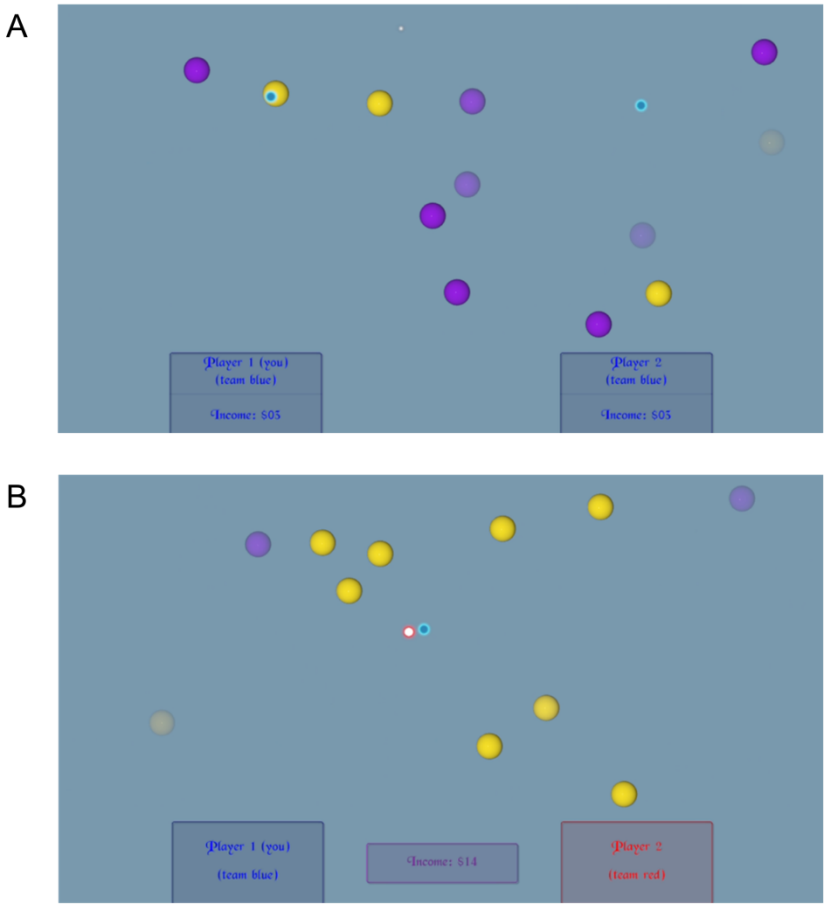
assigned to one of the two groups (Blue or Red) and were then led to believe that they were going to see photos of our research volunteers, who were also assigned to one of the groups in the past. Creating novel groups using a standard minimal group paradigm thus delineated between a group category participants belonged to (*ingroup*) and a group category participants did not belong to (*outgroup*).

Participants' expectations of future competition and cooperation were manipulated orthogonally to group category membership. Specifically, participants were told they would engage in competitive and cooperative shooting games with individual ingroup and outgroup members during the second part of the experiment. The instructions included descriptions of the two available versions of the shooting game followed by video examples (see Figure 2.2). Participants were told that in the competitive game they would be pitted against a second player, who could be either an ingroup or an outgroup member. The goal of the competitive game would then be to shoot more randomly appearing yellow dots than the second player, as only the player with a higher score (i.e., more shot yellow dots) at the end of a 30-second game would win an additional monetary prize. In the cooperative game, participants would supposedly collaborate with a second player, who could be either an ingroup or an outgroup member, to shoot as many purple dots randomly appearing on the screen within a 30-second interval. The players' individually scored points (i.e., the number of shot purple dots) would be added up to form their pair score (e.g., if the participant shot 4 purple dots and the other player shot 4, their final pair score would have been 8). If this pair score passed a certain threshold, left unspecified in the instructions, both parties would receive an additional monetary prize. Two details should be noted. First, as competition in the context of games is likely to be perceived as positive and fun, at least by some, I made sure that it is clear that competition in the shooting game has consequences in the form of one player being guaranteed to lose out on the bonus payment. Second, participants were led to believe that the

shooting game they would engage in with the second player did not depend on their own or the second player's choices but would instead be randomly determined. This was done to avoid implying varying degrees of trait competitiveness among targets and to define cooperation and competition as features of the social context rather than group members themselves. All of these instructions were included solely for the purpose of manipulating perceptions of competition or cooperation, and the second part of the experiment in fact never took place. However, to make the instructions and the whole two-part experiment set-up more believable, participants were also asked to precisely indicate their availability within the following week so that they could be adequately paired with available research volunteers. Finally, after the instructions, participants completed three manipulation check questions assessing the extent to which they identified with the ingroup and the outgroup (e.g., *I think I will like members of the Blue/Red team*; Vollberg & Cikara, 2022).

Figure 2.2

The shooting game used to induce expectations of cooperation and competition



Notes. (A) Example screenshot from a video depicting a competitive version of the shooting game with an ingroup member. (B) Example screenshot from a video depicting a cooperative version of the shooting game with an outgroup member. As part of instructions, participants watched four videos depicting the four possible combinations: (a) competitive game with an ingroup member, (b) cooperative game with an ingroup member, (c) competitive game with an outgroup member, and (d) cooperative game with an outgroup member.

Participants then proceeded to the emotion task (see Figure 2.3) which relied on the preceding cover story. Before starting the task, participants read the following introduction:

[Screen 1] Even brief exposure to other people can influence how we behave towards them later. Thus, before you interact with anyone in the second part of the experiment, we are going to introduce you to our research volunteers who were also assigned to either Team Blue or Team Red. Before you see each person, you will find out (a) which team they were

assigned to and (b) which version of the shooting game they will play with you in the second part of the experiment if you are paired. After this information, you will see how the person looks, and their neutral expression will then change to an emotional expression. We would like you to tell us how you think each person was feeling at the time of making their emotional expression, using a few slider scales.

[Screen 2] Reminder: The people in the images are real volunteers who were previously assigned to one of the two teams just like you and who both sent us their images and came to our lab to have their images taken. Their team membership and the shooting game they will play with you if paired will be indicated before each image.

After connecting the cover story with the emotion task, participants were shown how to use and interact with the valence, arousal, and genuineness slider scales and completed a single practice trial. Valence, arousal, and genuineness were defined as follows (Lazerus et al., 2016; Miller et al., 2022):

WHAT IS VALENCE? Valence has to do with how positive/pleasant or negative/unpleasant you think the person felt at the time of making the emotional expression. If you do not think the person's emotional expression communicates either positive/pleasant or negative/unpleasant feelings, but is neutral, you can choose the mid-point of the slider scale.

WHAT IS AROUSAL? Arousal has to do with how wide awake, alert or activated you think the person felt at the time of making the emotional expression, independent of whether you think their feelings were positive or negative. Combining valence and arousal, examples of emotions with positive valence and high arousal are ecstasy and excitement while feelings of calm and serenity are examples of emotions with positive valence and low arousal. Finally, feelings of rage and stress are examples of emotions with negative valence and high

arousal, while feelings of melancholy and sadness are examples of emotions with negative valence and low arousal.

WHAT IS GENUINENESS? Genuineness has to do with how much you think the person's emotional expression truly reflects what they were feeling inside at the time of making it. Sometimes facial expressions reflect genuinely felt emotion, but other times expressions are faked or posed (e.g., to be polite). An example of a genuine expression is when somebody smiles and they feel really happy, like when they get a present or see something funny. An example of a faked expression is when somebody smiles for a school photo, without feeling any emotion. Some expressions you will see were truly induced by exposing research volunteers to video clips, pictures or smells, or by having them "relive" an event. On the contrary, some expressions you are going to see are posed - people made them when they were instructed to act different emotions.

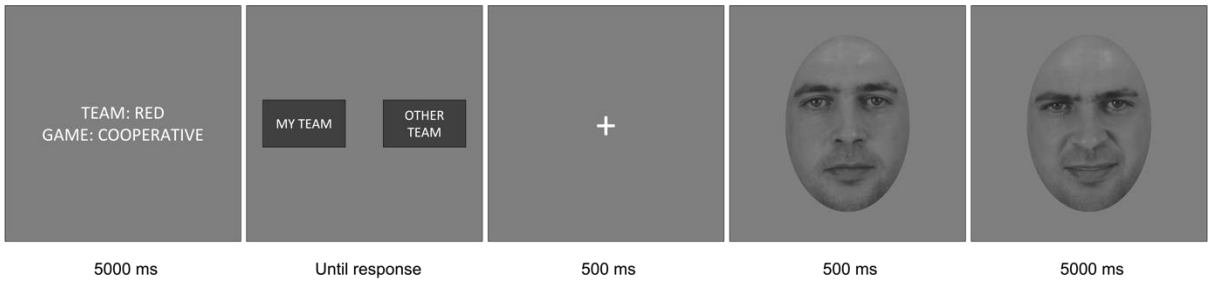
After completing a single practice trial, participants proceeded to the main section of the emotion task. In each trial, participants were first presented with a textual cue (5000 ms) indicating group category membership of the research volunteer in the image to follow and the version of the shooting game participants should expect to play with this research volunteer if they were paired in the second part of the experiment. This was immediately followed by two attention check questions (group attention check: *Does the person you are about to see belong to your team or the other team?*; coalitional relations attention check: *If you are paired with this person in the second part of the experiment, which version of the game will they play with you?*) where participants had to confirm the information contained in the textual cue by clicking on the appropriate buttons (i.e., either “my team” or “other team” for the first question, and either “competitive” or “cooperative” for the second question). No feedback was provided to participants on a trial-by-trial basis, but they were informed in the instructions that 80% of their attention check responses needed to be accurate

for their data to be included in the analysis. As mentioned above, two Prolific participants did not meet this requirement and were thus excluded. Furthermore, the individual trials where other participants made a mistake were also removed from the analysis. Having answered both attention check questions, participants briefly saw a fixation cross (500 ms), positioned to correspond to the tip of the target's nose, followed by the target's neutral facial expression (500 ms) which then changed to an ambiguous, angry, or happy facial expression (5000 ms). After the image was gone from the screen, participants were asked to rate valence, arousal, and genuineness of the target's expression on 100-point slider scales. The order of scales was counter-balanced across participants.

Having finished the emotion task, participants completed additional self-report questionnaires and demographic questions in random order. Specifically, to control for variables previously shown to be implicated in emotion judgments, the Toronto Alexithymia Scale (TAS-20, Parker et al., 2003) and Depression Anxiety Stress Scale (DASS-21, Lovibond & Lovibond, 1995) were administered to all participants. Furthermore, participants also indicated their intra- and intergroup game (i.e., coalitional) preference (*We would like to know which game you would prefer to play with your/other team if you had an option to choose.*), what they thought the first part of the experiment was about as well as to what extent they were convinced by different aspects of the cover story (e.g., existence of the second part of the experiment as well as real research volunteers). Participants also reported their gender, age, education, ethnicity, and first language. Finally, participants were debriefed and compensated.

Figure 2.3

The emotion task (online procedure)



Notes. The textual cue indicates that an upcoming person is a member of Team Red (i.e., an ingroup member if the participant is also a member of Team Read or an outgroup member if the participant is a member of Team Blue) with whom the participant is to cooperate if paired in the second part of the experiment. For brevity, the second attention check is not included in the figure.

Lab-based procedure. The lab experiment was also framed as if it included two parts, with participants' lab visit constituting the first part of the experiment and the second part scheduled to take place within the week following participants' lab visit. When participants signed up for the (first part of the) experiment, they were contacted by the researcher who asked them to fill in a short questionnaire before their arrival to the lab as well as to confirm when they would be available to complete the second part of the experiment online. In the pre-lab-visit questionnaire, participants were assigned to one of the two groups (Team Blue or Team Red) and read the instructions about different versions of the shooting game which were identical to those included in the online experiment. This pre-lab-visit questionnaire also included the TAS-20 (Parker et al., 2003), the DASS-21 (Lovibond & Lovibond, 1995), questions about participants' own intra- and intergroup game (i.e., coalitional) preferences as well as demographic questions and questions about participants' eyes (e.g., eye colour) and vision (e.g., vision health problems and vision correction).

Upon their arrival to the lab for the supposed first part of the experiment, participants completed a short reminder questionnaire which walked them through the main aspects of the cover story. Specifically, participants were reminded of their minimal group membership and

the available shooting game versions, after which they completed the same three manipulation check questions assessing the extent to which they identified with the ingroup and the outgroup (Vollberg & Cikara, 2022).

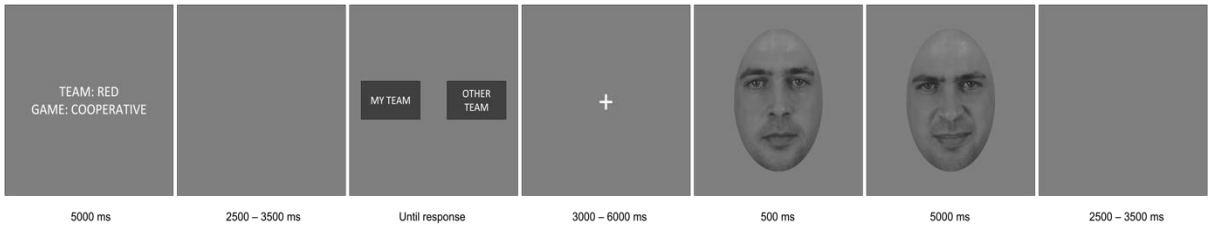
Having finished the reminder questionnaire, participants were escorted to the main testing room where they completed the emotion task while their gaze, pupil size and skin conductance response were monitored using the Gazepoint GP3-HD eye-tracker and the BIOPAC MP160 system. More details about the apparatus will be provided under Experiment 3, which focuses only on lab-based participants, their gaze patterns, and physiological responses. The emotion task was administered in PsychoPy (version 3.2.4, Peirce, 2007) and was identical to the emotion task distributed online, with the only difference being the duration of inter-trial intervals and inter-segment intervals within the same trial (see Figure 2.4). Specifically, the textual cue (5000 ms) introducing the group category membership of and the coalitional relations with the upcoming target was followed by the first inter-segment interval which was always between 2500 and 3500 ms long, but whose exact duration changed on a trial-by-trial basis. Following this inter-segment interval, lab-based participants completed the same attention check questions as did online participants (group attention check: *Does the person you are about to see belong to your team or the other team?*; coalitional relations attention check: *If you are paired with this person in the second part of the experiment, which version of the game will they play with you?*) after which they were presented with a fixation cross for anywhere between 3000 and 6000 ms, as opposed to online participants who only saw the fixation cross for 500 ms. The exact position of the fixation cross again varied on a trial-by-trial basis to correspond to the tip of the upcoming target's nose. After the fixation cross was gone from the screen, participants were presented with a neutral facial expression (500 ms) which then changed to an ambiguous, angry, or happy facial expression (5000 ms). The emotional facial expressions and valence,

arousal, and genuineness questions were separated by another inter-segment interval which could be anywhere between 2500 and 3500 ms long. Finally, after participants provided all three responses, there was an inter-trial interval, which could be anywhere between 3000 and 4500 ms long, before the textual cue for the next trial appeared. These inter-segment and inter-trial intervals were not in place in the online version of Experiment 1 as they are not directly relevant for participants' explicit judgments of facial expressions and would unnecessarily prolong the duration of the experiment. In the lab, however, I not only recorded participants' explicit judgments, but also their skin conductance response which normally take multiple seconds to develop (Kuhn et al., 2022). Thus, it was crucial to allow sufficient time before and after trial segments of interest – textual cues and facial expressions – so that I could be confident that any detected skin conductance response stemmed from the trial segment of interest, and not anything that preceded or followed this segment. While the inter-segment and inter-trial intervals were necessary to disambiguate the trigger of skin conductance response, I recognise that the prolonged time (3000-6000 ms as opposed to 500 ms) between attention checks, which are the final point in each trial where participants had to reiterate the group category membership of, and the coalitional relations with, the target, could have weakened the impact of the experimental manipulations on participants' judgments of facial expressions of emotion. This was, however, not confirmed as the analyses revealed the same pattern of results across online and lab-based participants. Therefore, online, and lab-based data were combined.

Having finished the emotion task, participants completed an exit questionnaire. This questionnaire asked participants to guess what the first part of the experiment was about and to indicate how convinced they were by different aspects of the cover story (e.g., existence of the second part of the experiment as well as real research volunteers). Finally, participants were debriefed and compensated.

Figure 2.4

The emotion task (lab-based procedure)



Differences between lab-based and online procedures. To reiterate, online participants completed the experiment in one go, starting with the cover story followed by the emotion task and self-report questionnaires (i.e., TAS-20, DASS-21, demographic questions) at the end. Each trial of the emotion task started with a textual cue which reflected the experimental condition (e.g., ingroup cooperation). This textual cue was immediately followed by two attention checks, a fixation cross, neutral, and emotional facial expressions, after which participants provided valence, arousal, and genuineness judgments of the preceding emotional facial expression. Lab-based participants, on the other hand, were exposed to the cover story and completed most self-report questionnaires prior to coming to the lab to complete the experiment (the time gap between completing the pre-lab-visit questionnaire and coming to the lab was, on average, 3 days). Upon their arrival to the lab, participants received a detailed reminder of the cover story, and proceeded to complete the emotion task. The emotion task used in the lab was structured in the same way as the emotion task used online, with the only difference being that trials in the lab-based emotion task lasted longer due to prolonged inter-trial and inter-segment intervals within the same trial (e.g., textual cues and attention checks were separated by a 2500-3500 ms inter-segment interval in the lab, while they were not separated online). The differences between lab-based and online procedures are emphasised for transparency, but it should be noted that statistical analyses

revealed the same pattern of results for both online and lab-based participants, which is why these differences will not be discussed further.

Analysis strategy. Participants' responses were analysed using linear mixed models implemented in R (version 4.1.3) package lme4 (version 1.1.31 – Bates et al., 2014), and lmerTest (version 3.1.3 – Kuznetsova et al., 2017). Specifically, the fixed effects of emotion category (levels: ambiguous, angry, happy) group category membership (levels: ingroup member, outgroup member), and coalitional relations (levels: competitive, cooperative), on valence, arousal, and genuineness judgments of facial expressions of emotion were tested. In addition to the fixed effects of interest, I included random intercepts for participants and targets' identities and a correlated random slope for emotion within participant[23]. This random structure was determined by fitting maximal random models (Barr et al., 2013), and then resolving convergence and singularity warnings by simplifying the random structure using Principal Component Analysis to determine the most relevant random components (Bates et al., 2015). This random structure was selected prior to adding any fixed effects to the model (Meteyard & Davies, 2020).

[23] Trial was not included as either a random intercept or a random slope on participants for the following reasons. First, specifying trial as an intercept would assume that judgments on one trial are independent from judgments on another trial. This assumption is violated given that judgments of the same target's facial expressions are likely to be more similar to each other than jugdments of different targets' facial expressions, making targets' identities a more suitable random component. Second, specifying trial as a random slope on participants would allow each participant to display a unique decrease or increase in their judgments over the couse of the task. However, based on the visualisations of each participant's trial-by-trial judgments, there was no reason to include a random slope of trial on participants because individual particpants' slopes were very similar and almost completely flat. Furthermore, including the random slope of trial on participants, especially since trial is not simultaneously a fixed effect, would likely make the model unnecessarily complex and unparsimonious, going against what has been recommended by leading proponents and developers of linear mixed models (Bates et al., 2015).

Results

Valence judgments. Unsurprisingly, the main effect of emotion category was significant ($F(2, 166.5) = 985.45$, $p < .001$). Specifically, participants rated happy expressions as conveying the most positive emotions ($M = 75.02$, $SD = 14.67$), followed by ambiguous ($M = 43.63$, $SD = 15.40$) and angry expressions ($M = 32.24$, $SD = 13.57$) with all Holm-corrected pairwise comparisons statistically significant at $p < .001$. Furthermore, there was a significant main effect of targets' group category membership ($F(1, 15345.1) = 11.98$, $p < .001$) indicating that ingroup members' expressions were rated as conveying more positive emotions ($M = 50.69$, $SD = 23.17$) in comparison to outgroup members' expressions (*ingroup positivity bias*, $M = 50.01$, $SD = 23.34$). The main effect of coalitional relations was also statistically significant ($F(1, 15348.8) = 6.46$, $p = .011$) such that expressions of targets with whom participants expected to cooperate were judged as conveying more positive emotions ($M = 50.56$, $SD = 23.33$) than expressions of targets with whom participants expected to compete (*cooperation positivity bias*, $M = 50.13$, $SD = 23.18$). The interaction between coalitional relations and emotion category did not reach conventional statistical significance ($F(2, 15321.7) = 2.85$, $p = .058$) but the trend suggested that cooperation positivity bias was observed primarily when participants judged valence of ambiguous expressions ($b = 0.74$, 95% CI = [0.09, 1.38], $z = 2.25$, $p = .025$) and happy expressions ($b = 0.87$, 95% CI = [0.22, 1.51], $z = 2.63$, $p = .008$). When participants judged valence of angry expressions, no cooperation positivity bias was observed ($b = -0.15$, 95% CI = [-0.80, 0.50], $z = 0.46$, $p = .645$). No other effects were observed (all $ps > .141$).

Arousal judgments. The main effect of emotion category was significant ($F(1, 166.6) = 130.18$, $p < .001$). Specifically, participants evaluated happy expressions as conveying the highest arousal ($M = 64.91$, $SD = 20.22$), followed by angry ($M = 53.72$, $SD = 21.25$) and ambiguous expressions ($M = 46.58$, $SD = 19.86$) with all Holm-corrected pairwise

comparisons statistically significant at $p < .001$. Other main and interaction effects were not observed (all *ps* > .079).

Genuineness judgments. The main effect of emotion category was again significant ($F(1, 166.7) = 180.87$, $p < .001$). Specifically, participants evaluated happy expressions as the most genuine ($M = 69.68$, $SD = 23.09$), followed by angry ($M = 50.44$, $SD = 23.69$) and ambiguous expressions ($M = 36.74$, $SD = 23.62$) with all Holm-corrected pairwise comparisons statistically significant at $p < .001$. No other effects were observed (all *ps* > .118).

Collective identification. As a manipulation check, participants filled in a collective identification scale (Vollberg & Cikara, 2022). Ingroup and outgroup identification scores were calculated by summing up the items referring to each group, with possible scores ranging from 3 to 18, where a higher value indicates stronger collective identification ($\alpha_{INGROUP} = .87$, $\alpha_{OUTGROUP} = .91$). The scores were then submitted to a one-way ANOVA, which showed that participants identified significantly more with their ingroup ($M = 12.35$, $SD = 4.02$) than with the outgroup ($M = 4.92$, $SD = 3.03$, $F(1, 167) = 391.05$, $p < .001$, $d = 1.53$, 95% CI [1.29, 1.77]). While this scale is unlikely to reflect any genuine and long-lasting identification with minimal groups, which is why it is primarily treated as a manipulation check, how much participants discriminate in their collective identification responses can be taken as an indicator of their responsivity to newly created group divisions (see Kranton et al., 2020). As such, the degree of discrimination in participants' collective identification responses, calculated by subtracting outgroup identification scores from ingroup identification scores, was included in an exploratory analysis to test whether the influence of target's group category membership on explicit judgments of facial expressions is moderated by participants' responsivity to group divisions. The inclusion of this exploratory variable did not improve model fit in the domain of valence ($\Delta AIC = 4$, $p = .830$), arousal ($\Delta AIC = 1$, $p =$

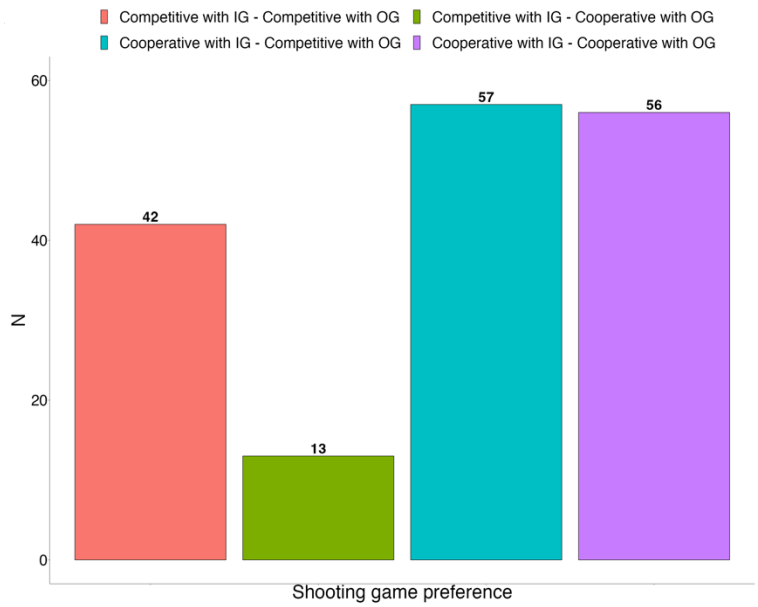
.289) or genuineness (ΔAIC = 3, p = .501), and I observed no trends of it interacting with targets' group category membership (all ps > .322).

Intra- and intergroup coalitional preferences. Even though participants believed that the version of the shooting game – competitive or cooperative – they would end up playing with others in the second part of the experiment did not depend on their or other players' choices, they were still asked which version of the shooting game they would prefer to play with minimal ingroup and outgroup members if they had an option to choose. From here, I tested the prediction that the second player's group category membership would make a difference in participants' own shooting game preferences. Specifically, I fitted a generalised linear mixed model which included the second player's group category membership (levels: ingroup member, outgroup member) as a fixed factor, an intercept for participants as a random factor and participants' shooting game preference (0 = cooperative game preferred, 1 = competitive game preferred) as the dependent variable. Please note that the output of generalised linear mixed models relies on odds – *P(competitive game preferred) / P(cooperative game preferred)* – rather than probabilities – *P(competitive game preferred)*. Thus, the odds of preferring a competitive shooting game increased by 292% if the second player was an outgroup member in comparison to an ingroup member (log-odds output: estimate = 1.37[24], SE = 0.29, z = 4.66, p < .001, see Figure 2.5).

Figure 2.5

Intra- and intergroup coalitional preferences (Experiment 1)

[24] The estimate of 1.37 means that in comparison to an ingroup player, an outgroup player is associated with a 1.37 increase in the log-odds of participants choosing a competitive game compared to a cooperative game. Since log-odds are difficult to interpret, they are commonly exponentiated to derive an odds ratio. Exponentiating a log-odds value of 1.37 results in an odds ratio of 3.92 which indicates a 292% increase in the odds of preferring a competitive shooting game if the second player was an outgroup in comparison to an ingroup member.



Notes. The number of participants who made one of the four choice combinations. IG = Ingroup. OG = Outgroup.

Like the scores representing differences between reported ingroup and outgroup identification, whether participants discriminated in their shooting game preferences on the basis of second player's group membership can also be understood as an indicator of their responsivity to group divisions (Kranton et al., 2020). Thus, in another exploratory analysis, I tested whether the influence of target's group category membership on explicit judgments of facial expressions was moderated by this indicator of participants' responsivity to group divisions. For simplicity, responsivity to group divisions, operationalised as a binary variable which captured whether participants discriminated in their shooting game preferences in favour of the ingroup (N = 57) or their preferences were unaffected by second player's group category membership (N = 98), was included in the original linear mixed models as part of a two-way interaction with targets' group category membership. Participants who discriminated in favour of the outgroup (N = 13) were excluded from this analysis as the motivation behind their responses was elusive but still highly likely different from the motivation of participants who discriminated in favour of the ingroup. The inclusion of this exploratory variable did not improve model fit in the domain of arousal ($\Delta AIC = 6$, $p = .943$)

or genuineness (ΔAIC = 5, p = .884), nor did it display trends of interacting with targets' group membership (all ps > .668). However, the model with this exploratory variable fit participants' valence judgments significantly better than the simpler model without it (ΔAIC = -2, p = .041) and revealed a significant two-way interaction between participants' responsiveness to group divisions and targets' group category membership (F(1, 14525.3) = 6.16, p = .013)[25]. Specifically, participants who discriminated in their coalitional preferences in favour of the ingroup also displayed significantly larger ingroup positivity bias in their judgments of facial expressions of emotion (b = 1.35, 95% CI = [0.64, 2.07], z = 3.70, p < .001) than participants who did not discriminate in their coalitional preferences (b = 0.21, 95% CI = [-0.35, 0.76], z = 0.73, p = .469)[26]. This finding implies that there might be individual differences in how much people's judgments, preferences and behaviours are affected by group category membership, such that those who are sensitive to group category membership in one domain (e.g., coalitional preferences) are more likely to be sensitive to it in a different domain (e.g., valence judgments of facial expressions).

Control measures. Descriptive statistics for other self-report questionnaires and convincingness questions are included in Table 2.2. Please note that controlling for any of these measures in the analysis did not change the results concerning participants' judgments of facial expressions of emotion. Furthermore, I examined whether the extent to which

[25] To ascertain whether greater discrimination in favour of the ingroup on one measure (e.g., coalitional preferences) is related to greater discrimination in favour of the ingroup on another measure (e.g., valence judgments of facial expressions), I was primarily interested in the two-way group category membership x responsiveness interaction. However, for transparency, please note that the main effect of participants' responsiveness to group divisions was not significant (F(1, 152.1) = 0.24, p = .626).

[26] In alternative exploratory models which allowed participants' responsiveness to group divisions to interact with coalitional relations, no such interaction was ever observed. This confirmed that ingroup-favouring discrimination in coalitional preferences predicted only the size of ingroup positivity bias, but not the size of cooperation positivity bias.

participants were convinced by different aspects of the experimental manipulation moderated any of the reported effects. These analyses revealed that valence, arousal, and genuineness judgments of different categories of facial expressions (angry, ambiguous, happy) varied between participants who were convinced by the experimental manipulation and those who were not. For example, participants who were more convinced that people they were seeing were real research volunteers rated angry expressions as conveying more negative emotions (b = -1.27, 95% CI = [-2.27, -0.27], z = -2.51, p = .010) and happy expressions as conveying more positive emotions (b = 1.11, 95% CI = [0.11, 2.11], t = 2.20, p = .030) than participants who were less convinced. Importantly, no interactions that involved either group category membership (all *ps* > .085) or coalitional relations (all *ps* > .067) were observed.

Table 2.2

Descriptive statistics of included self-report questionnaires and convincingness questions (Experiment 1)

	M	Mdn	SD	MAD	Min	Max
Convincingness (Shooting game)	4.95	5	1.39	1.48	0	6
Convincingness (Research volunteers)	3.91	4	1.59	1.48	0	6
DASS-21 (Anxiety)	6.04	4	6.49	5.93	0	36
DASS-21 (Depression)	9.75	8	8.75	8.90	0	42
DASS-21 (Stress)	11.70	10	8.55	8.90	0	42
TAS-20	49.94	44	11.73	11.86	22	81

Notes. The first convincingness question was phrased as follows: "To what extent were you convinced that there is an online game where you would interact with other research volunteers?". The second convincingness question was phrased as follows: "To what extent were you convinced that the people in the images were other research volunteers?". Possible convincingness responses ranged from 0 to 6. Possible DASS-21 scores ranged from 0 to 42 for all three subscales. Possible TAS-20 scores ranged from 20 to 100.

Discussion

Experiment 1 showed that valence judgments of angry, ambiguous, and happy facial expressions of emotion are affected by group category membership of the expresser (i.e., whether the expresser is an ingroup or an outgroup member) and coalitional relations with the expresser (i.e., whether one expects to interact with the expresser in a cooperative or a competitive manner), while it found no evidence that these two factors influence arousal and genuineness judgments of the same facial expressions. Specifically, participants in Experiment 1 judged minimal ingroup members' facial expressions of emotion as conveying more positive affect compared with minimal outgroup members' facial expressions (i.e., *ingroup positivity bias*). Additionally, coalitional relations made a difference to valence judgments as participants reported that emotional expressions of people with whom they expected to cooperate communicated more positive emotions than expressions of people with whom they expected to compete (i.e., *cooperation positivity bias*). These results have significant theoretical implications. Notably, in a research context where many "ingroup" biases documented in minimal intergroup environments have been recast as "cooperation" biases (Misch et al., 2021), Experiment 1 provides the first evidence that ingroup bias in valence judgments of facial expressions documented by Lazerus and colleagues (2016) cannot be fully accounted for by underlying coalitional relations. In other words, while the work of Lazerus and colleagues (2016) made it impossible to parse whether "ingroup" bias in valence judgments of facial expressions indeed stemmed from expressers' group category membership or rather co-occurring coalitional relations, Experiment 1, which treated group category membership and coalitional relations as orthogonal, demonstrated that both factors are likely to have independently contributed to Lazerus and colleagues' (2016) original results. Nonetheless, participants' own coalitional preferences, wherein cooperation was more frequently chosen if the second shooting game player was an ingroup as opposed to an

outgroup member, underscore that group category membership and coalitional relations are correlated in real life, such that cooperation frequently co-occurs with sharing a group category membership.

Interestingly, exploratory analyses further revealed that participants who engaged in group-related discrimination on an unrelated measure of coalitional preferences showed stronger ingroup positivity bias in comparison to participants who did not discriminate on the basis of group category membership. This finding implies that there might be individual differences in how much people's judgments, preferences and behaviours are affected by group category membership and parallels a recently proposed shift in how minimal groups are thought of and used (Kranton et al., 2020). Under this shift, the minimal group paradigm is explicitly conceptualised as an individual differences-paradigm, allowing delineation between those whose social preferences are consistently shaped by available group category membership ("groupy" individuals) and those whose social preferences are mostly left unaffected by available group category membership ("non-groupy" individuals). Despite opening a discussion on the value of construing the minimal group paradigm as an individual differences-paradigm, this finding should not be overinterpreted in the context of current research due to its exploratory nature. It is also important to note that the same "groupiness-diagnostic" consistency was not observed for group-based discrimination in the domain of collective identification. That is, how much participants discriminated between the minimal ingroup and the minimal outgroup in their collective identification responses did not similarly predict how much targets' group category membership influenced participants' valence judgments of facial expressions. While this might suggest that different ingroup-favouring preferences are not equally diagnostic of individuals' general "groupiness", it nonetheless remains a possibility that the collective identification measure, which was administered immediately after the group assignment and the cover story, seemed to participants like an

attention check and was thus more prone to demand characteristics than the measure of coalitional preferences.

Experiment 2

The results of Experiment 1 provided initial support for an *ingroup positivity bias* and a *cooperation positivity bias* in judgments of facial expressions of emotion. However, due to the inclusion of angry, ambiguous, and happy facial expressions in the emotion task, the most obvious and salient differences between the emotional expressions concerned their positions on a happiness-to-anger continuum or more broadly, on a positive-to-negative continuum. This positioning of emotion categories could have easily been recognised by participants in a matter of a few trials, restricting their valence judgments to lower, middle, and upper sections of the slider scale corresponding to angry, ambiguous, and happy expressions respectively. To illustrate, in the case of ambiguous facial expressions, as the task progressed, participants could have directly compared expressions with each other and realised that ambiguous expressions were the "in between" ones, which could have consequentially restricted the variability of valence judgments and affected our statistical inferences (Ciarleglio, 2021). Thus, in Experiment 2, I wanted to address this limitation and test if the group- and coalition-effects replicate in the ambiguous condition only, where comparison with other obviously different expressions is not possible. Furthermore, online participants, who constituted a majority of the sample in Experiment 1 (N = 132), reported their coalitional (i.e., shooting game) preferences with ingroup and outgroup members at the very end of the experiment. This leaves open the possibility that their coalitional preferences were biased by responses to previously presented ingroup and outgroup members, facial expressions and/or anticipated shooting game versions. For instance, participants who would not normally prefer to compete with ingroup members could be significantly more open to the possibility of within-group competition after realising that group category membership and coalitional relations are

treated as orthogonal in the experiment. Therefore, the order of tasks in Experiment 2 was also changed so that participants indicated their coalitional preferences immediately after being introduced to different versions of the shooting game and prior to completing the emotion task.

Method

Participants. Using the estimates from the ambiguous condition in Experiment 1, a simulated power analysis (DeBruine & Barr, 2019) revealed that to obtain the same pattern of results found in Experiment 1 with sufficient statistical power, a sample of about 80 participants and 56 ambiguous stimuli should suffice. As in Experiment 1, eligible participants had to reside in the UK, be at least 18 years old, have no formal education in psychology, have no history of diagnosed mental health disorders or autism and have normal colour vision. A total of 95 participants (52.6% men, 47.4% women) between the age of 19 and 43 (M_{AGE} = 23.80; SD_{AGE} = 4.68) were recruited to account for potential exclusions. Participant recruitment was conducted on Prolific in June 2020 (www.prolific.co). Due to a glitch in the Gorilla Experiment Builder's (Anwyl-Irvine et al., 2020) server, emotion task responses of three participants were not recorded, while their pre-task and post-task responses were adequately stored. No participants were otherwise excluded.

Facial stimuli. Previously validated ambiguous expressions (N = 56; $M_{VALENCE}$ = 38-61.45; $Median_{VALENCE}$ = 40-59; $SD_{VALENCE}$ = 6.55-22.15; $M_{AROUSAL}$ = 21.10-64.82; $Median_{AROUSAL}$ = 23.50-66; $SD_{AROUSAL}$ = 8.99-26.53; $M_{GENUINENESS}$ = 28.64-73.84; $Median_{GENUINENESS}$ = 25.00-77.50; $SD_{GENUINENESS}$ = 13.48-32.54[27]), created by morphing angry and happy expressions of targets from the CFD (Ma et al., 2015) and RFD (Langner et al., 2010), were used as stimuli.

[27] Mean, median and standard deviation values are from the pilot study.

Procedure. The procedure was almost identical to the online procedure of Experiment 1. The order of tasks was changed so that participants indicated their shooting game (i.e., coalitional) preferences prior to completing the emotion task. This was done to avoid any possibility of participants' preferences being biased by responses to previously presented ingroup and outgroup members, their facial expressions and/or anticipated shooting game versions. Furthermore, in Experiment 1, I did not have any insight into participants' motivation behind choosing same or different shooting game versions to play with the ingroup or outgroup members. For this reason, in Experiment 2, I also included two open-ended questions asking participants to give reasons for their within-group and between-group coalitional preferences.

Analysis strategy. Participants' judgments of facial expressions were again analysed using linear mixed models. Specifically, the fixed effects of targets' group category membership (levels: ingroup member, outgroup member), and coalitional relations (levels: competitive, cooperative) on valence, arousal, and genuineness judgments of ambiguous facial expressions of emotion were tested. In addition to the fixed effects of interest, I included random intercepts for participants and targets' identities. This random structure was determined in the same manner as in Experiment 1.

Results

Valence judgments. I replicated the results from Experiment 1. Specifically, I observed a main effect of targets' group category membership ($F(1, 4824.3) = 15.15$, $p < .001$) with ingroup members' ambiguous expressions ($M = 50.70$, $SD = 19.73$) communicating more positive emotions than outgroup members' expressions ($M = 48.42$, $SD = 19.44$). I also observed a main effect of coalitional relations ($F(1, 4826.0) = 10.88$, $p < .001$) with expressions of targets with whom participants expected to cooperate ($M = 50.14$, $SD = 19.50$) conveying more positive emotions than targets with whom participants expected

to compete (M = 48.99, SD = 19.71). There was no significant two-way interaction (F(1, 4825.4) = 0.18, p = .667).

Arousal judgments. In line with Experiment 1, the main effect of targets' group category membership (F(1, 4833) = 0.43, p = .514), coalitional relations (F(1, 4836.4) = 2.05, p = .153), as well as the interaction effect (F(1, 4835.8) = 0.37, p = .545) were not significant.

Genuineness judgments. As in Experiment 1, I observed no significant effects of targets' group category membership (F(1, 4842.2) = 1.02, p = .314) and coalitional relations (F(1, 4847) = 1.48, p = .224), nor a significant two-way interaction (F(1, 4846.3) = 0.004, p = .953).

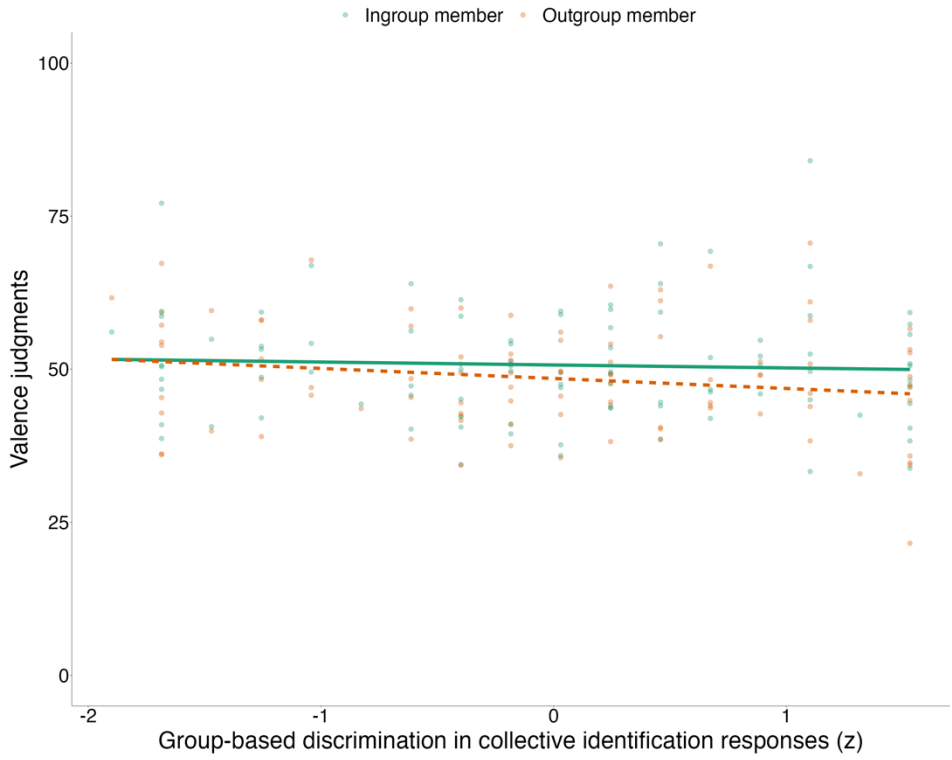
Collective identification. Ingroup and outgroup identification scores were calculated identically to Experiment 1 ($\alpha_{INGROUP}$ = .88, $\alpha_{OUTGROUP}$ = .95). One-way ANOVA again showed that participants identified more strongly with the ingroup (M = 11.94, SD = 4.38) than with the outgroup (M = 4.01, SD = 2.26) (F(1, 94) = 259.20, p < .001, d = 2.27, 95% CI [1.91, 2.64]). As in Experiment 1, I examined whether participants who discriminated more in their collective identification responses were also those who were affected more by targets' group category membership in their judgments of ambiguous facial expressions. Including participants' group-based discrimination in collective identification, calculated as a difference score between ingroup and outgroup identification, did not improve model fit in the domain of arousal (ΔAIC = 1, p = .190) or genuineness (ΔAIC = 4, p = .789). In the domain of valence, however, I observed a marginal two-way interaction between targets' group membership and participants' collective identification difference score (F(1, 4822.1) = 3.95, p = .047, model improvement: ΔAIC = -2, p = .067)[28]. This interaction suggested that participants who discriminated more in their collective identification responses were those

[28] The main effect of collective identification difference score was not significant (F(1, 89.9) = 1.45, p = .231).

who were more affected by targets' group category membership in their valence judgments of facial expressions (see Figure 2.6).

Figure 2.6

The interactive effect of participants' group-based discrimination in collective identification responses and targets' group membership on valence judgments of facial expressions



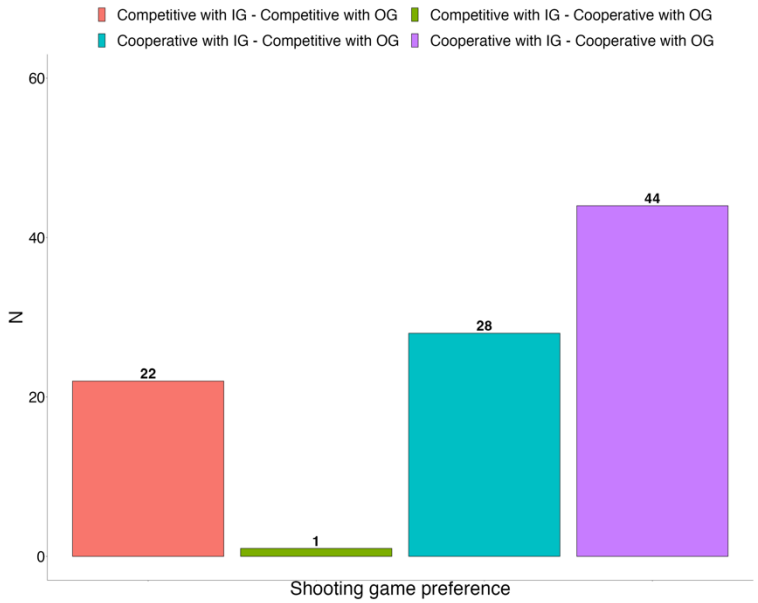
Notes. Neither of the simple slopes was significant (ingroup member: estimate = -0.61, SE = 0.93, t = -0.66, p = .510; outgroup member: estimate = -1.55, SE = 0.93, t = -1.67, p = .100).

Intra- and intergroup coalitional preferences. Like in Experiment 1, I fitted a generalised linear mixed model with the second player's group membership (levels: ingroup member, outgroup member) as a fixed factor, an intercept for participants as a random factor and participants' shooting game preference (0 = cooperative game preferred, 1 = competitive game preferred) as the dependent variable. The odds of preferring a competitive shooting game increased by 1246% if the second player was an outgroup member in comparison to an ingroup member (log-odds output: estimate = 2.60, SE = 0.75, z = 3.47, p < .001, see Figure

2.7). The reasons for participants' intra- and intergroup coalitional preferences, reported in two open-ended questions, are not discussed in the current chapter.

Figure 2.7

Intra- and intergroup coalitional preferences (Experiment 2)



Notes. The number of participants who made one of the four choice combinations. IG = Ingroup. OG = Outgroup.

I again tested whether the influence of target's group category membership on judgments of facial expressions was stronger among participants who discriminated in their coalitional (i.e., shooting game) preferences. As in Experiment 1, responsivity to group divisions was operationalised as a binary variable which captured whether participants discriminated in their shooting game preferences in favour of the ingroup (N = 27) or their preferences were unaffected by second player's group membership (N = 65)[29]. This binary variable was included in the original linear mixed models as part of a two-way interaction with targets' group category membership. This model revealed a two-way interaction

[29] The numbers in the brackets differ from those in Figure 2.7, as emotion task data of three participants were lost due to a server glitch.

between participants' responsiveness to group divisions, captured by the presence of group-based discrimination in their coalitional preferences, and targets' group category membership ($F(1, 4822) = 6.26$, $p = .012$, model improvement: $\Delta AIC = -3$, $p = .038$)[30]. Specifically, participants who showed ingroup-favouring discrimination in their coalitional preferences displayed significantly larger ingroup positivity bias in their judgments of facial expressions of emotion ($b = 3.71$, 95% CI = [1.98, 5.44], $z = 4.21$, $p < .001$) than participants who did not discriminate in their coalitional preferences ($b = 1.09$, 95% CI = [-0.01, 2.20], $z = 1.94$, $p = .052$)[31]. Finally, the inclusion of the same variable did not similarly improve model fit in the domain of arousal ($\Delta AIC = 3$, $p = .415$) or genuineness ($\Delta AIC = 0$, $p = .089$), nor did it display trends of interacting with targets' group category membership (all $ps > .507$).

Control variables. Table 2.3 provides descriptive statistics for other self-report questionnaires and convincingness questions that were included in Experiment 2. Please note that controlling for any of these measures in the analysis did not change the results concerning participants' judgments of facial expressions of emotion. Furthermore, as in Experiment 1, I examined whether the extent to which participants were convinced by different aspects of the experimental manipulation moderated any of the reported effects, but no such moderation was observed (all $ps > .092$).

[30] The main effect of participants' responsiveness to group divisions was not significant ($F(1, 90.0) = 0.26$, $p = .609$).

[31] In alternative exploratory models which allowed participants' responsiveness to group divisions to interact with coalitional relations, no such interaction was ever observed.

Table 2.3

Descriptive statistics of included self-report questionnaires and convincingness questions (Experiment 2)

	M	Mdn	SD	MAD	Min	Max
Convincingness (Shooting game)	5.17	6	1.14	0	1	6
Convincingness (Research volunteers)	3.86	4	1.84	2.97	0	6
DASS-21 (Anxiety)	4.82	2	6.11	2.97	0	26
DASS-21 (Depression)	11.52	10	8.41	8.90	0	32
DASS-21 (Stress)	12.53	12	9.04	8.90	0	42
TAS-20	45.87	44	11.37	11.86	26	76

Notes. The first convincingness question was phrased as follows: "To what extent were you convinced that there is an online game where you would interact with other research volunteers?". The second convincingness question was phrased as follows: "To what extent were you convinced that the people in the images were other research volunteers?". Possible convincingness responses ranged from 0 to 6. Possible DASS-21 scores ranged from 0 to 42 for all three subscales. Possible TAS-20 scores ranged from 20 to 100.

Discussion

Experiment 2 fully replicated Experiment 1's findings using a larger set of ambiguous facial expressions. Specifically, participants again judged ingroup members' ambiguous expressions as conveying more positive feelings than outgroup members' ambiguous expressions after a minimal group manipulation (i.e., *ingroup positivity bias*), regardless of the coalitional relations between and within groups. Coalitional relations were also shown to impact valence judgments, with ambiguous expressions of people linked to a cooperative context being judged as more positive than expressions of people linked to a competitive context (i.e., *cooperation positivity bias*), regardless of their ingroup or outgroup status. Furthermore, participants' own coalitional preferences, wherein cooperation was more frequently chosen if the second shooting game player was an ingroup as opposed to an outgroup member, underscored that group category membership and coalitional relations frequently overlap in real life. Finally, the participants who showed a greater tendency to discriminate in their collective identification responses and coalitional preferences on the basis of group category membership were also the participants who were more affected by targets' group membership in their valence judgments of facial expressions. Thus, Experiment 2 reinforced the possibility that there might be meaningful individual differences in people's responsivity to group divisions, such that some people tend to display group-based discrimination in a variety or judgments, preferences, and behaviours while other people's judgments, preferences, and behaviours are less affected by group category membership.

Experiment 3

Experiments 1 and 2 consistently documented an *ingroup positivity bias* and a *cooperation positivity bias* in judgments of facial expressions of emotion leading to the conclusion that people are likely to reach different judgments of facial expressions depending

on the group category membership of, and coalitional relations with, the expresser. However, it remains an open question how exactly group category membership and coalitional relations exert their influence on valence judgments of facial expressions. One possibility could be that group category membership of, and coalitional relations with, the expresser change how people visually attend to their facial features. Another possibility could be that these factors come into play "post-perceptually", influencing only judgments of otherwise similarly explored facial expressions (see Lazerus et al., 2016). Thus, in Experiment 3, I sought to look under the hood of self-reported judgments, investigating whether group category membership and coalitional relations elicit changes in the patterns of visual attention paid to different features of facial expressions and whether any group- and coalition-induced attentional shifts can explain the corresponding biases in judgments of facial expressions. Furthermore, Experiment 3 employed two measures of physiological arousal – pupil dilation and skin conductance response – to index implicit processing of social information about the expresser. In particular, I explored whether congruent and incongruent social information about the expresser (e.g., congruent, and incongruent pairings of group category membership and coalitional relations) result in differential physiological responsivity. Given that Experiment 3 is the only experiment in this thesis which uses eye-tracking and peripheral physiology, I will start by offering an experiment-specific introduction, describing the methods and their potential to help us understand cognitive and affective mechanisms which underlie biases in intergroup social perception. I will also discuss past work which adopted related methods to study the influence of group category membership on face (i.e., identity) recognition and emotion judgments.

Eye-tracking and visual attention

Attempts to understand and track human gaze are not recent; the first descriptions of human eye movements can be found in the 19th century and the first eye-tracking device,

although highly intrusive and substantially different from the devices we are used to today, is thought to have been designed and described by Edmund Huey in 1908 (Huey, 1908). Modern eye-trackers are no longer intrusive but instead operate by directing infrared light at the eye from a distance causing reflections in both the cornea (i.e., the transparent front part of the eye) and the pupil (i.e., the black round part at the centre of the iris), which allow the device to identify the position of gaze as well as changes in pupil diameter (Rahal & Fiedler, 2019). This information can then be used to identify meaningful events such as *fixations*, gaze "pauses" during which informative regions of interest are attended to, and *saccades*, rapid eye movements between fixations (Salvucci & Goldberg, 2000). Given that Experiment 3, like most social psychological research using eye-tracking data, focuses on fixations, it is important to clarify how these gaze events are commonly interpreted in the literature.

In essence, the occurrence of fixations indicates that the stimulus that is fixated on is held in the foveal region, which is responsible for sharp central vision (van Diepen & d'Ydewalle, 2003). As such, fixations are commonly interpreted as indices of *visual attention*, a cognitive mechanism that allows selective processing of otherwise overwhelming visual input (Rahal & Fiedler, 2019). Whereas it was originally assumed that what the eyes fixate on is what the mind processes (i.e., *eye mind hypothesis*, Just & Carpenter, 1980), more recent accounts have suggested that the relationship between fixations and cognitive processing is better understood as probabilistic rather than deterministic (Rahal & Fiedler, 2019). In other words, visual attention, indexed by fixations, does not guarantee cognitive processing of specific content in working memory, but it does make it more probable in comparison to non-fixated content.

Bottom-up and top-down attentional shifts. How we deploy our visual attention is multidetermined (Carrasco, 2011; Pinto et al., 2013). On one hand, attentional shifts can be caused by characteristics of the sensory input (i.e., *bottom-up attentional shifts*) – for

example, if someone suddenly starts gasping for air in an otherwise quiet room, they are guaranteed to capture our fixations, and by extension, our visual attention. On the other hand, what we attend to in our surroundings depends on what our preferences and goals are in a specific situation (i.e., *top-down attentional shifts*) – for example, we will fixate on different parts of our room when we are searching for our mobile phone and when we are procrastinating while typing our thesis. Bottom-up and top-down attentional shifts are also at play when we are looking at others' faces – our attention can be driven by salient facial characteristics (e.g., eye colour, facial scars) and facial movement (e.g., development of a broad smile) as much as by our preferences and goals during the specific social interaction (e.g., whether we are trying to better understand someone with an accent unfamiliar to us or whether we are trying to determine if someone likes us). In relation to the current research, it is important to note that while real-world group membership can arguably trigger both bottom-up attentional shifts (e.g., due to pronounced facial features) and top-down attentional shifts (e.g., due to pre-existing associations), minimal group membership's influence is likely to be restricted to top-down attentional shifts especially if the target's minimal group membership is communicated ahead of the presentation of their face and if the facial stimuli have been randomly assigned to minimal groups, thus controlling for any systematic differences in low-level facial features.

The influence of group category membership on visual attention to faces. To date, group membership's influence on visual attention has mostly been examined in the context of race, and more specifically *own-race bias*, which refers to people's tendency to recognise own-race faces more accurately than other-race faces regardless of the racial groups in question (Meissner & Brigham, 2001). I will therefore use research investigating this bias as an illustration of how eye-tracking methodology can be used to examine attentional shifts to faces and specific facial features in intergroup contexts.

While own-race bias is one of the most consistently documented biases in face perception, there exists a considerable debate surrounding its causes. On one hand, the perceptual-expertise account argues that own-race bias is caused by members of different racial groups having insufficient exposure to each other's faces leading to suboptimal allocation of visual attention to diagnostic facial features of other-race faces which translates to poorer recognition of these faces relative to own-race faces (Chiroro & Valentine, 1995; Walker & Hewstone, 2006; Singh et al., 2022). On the other hand, the socio-cognitive account of own-race bias posits that people are more motivated to attend to individuating characteristics of own-race faces, whereas visual attention to other-race faces is characterised by greater focus on category-diagnostic features (Hugenberg et al., 2007; Hugenberg et al., 2010; Levin, 2001). Both accounts have support in the literature. Given that minimal groups do not differ in their diagnostic facial features and people are guaranteed not to have different levels of exposure to these groups, the empirical evidence generated in support of the perceptual-expertise account is unlikely to be relevant in a minimal group experiment. For this reason, I will focus on the socio-cognitive account and specifically, on evidence concerning visual attention to faces.

The socio-cognitive account of own-race bias has to date mostly been supported by non-attentional evidence: (a) experiments manipulating participants' individuation motives and documenting that these motives lead to reduction in own-race bias in face recognition (Young & Hugenberg, 2012), and (b) experiments employing a minimal group paradigm and documenting that lab-created novel groups result in similar *own-group bias* (Bernstein et al., 2007). Important for the discussion of visual attention to faces and its interplay with individuation motives, however, is the work of Kawakami and colleagues (2014). In this work, Kawakami et al. (2014) asked the following research questions (a) whether White participants preferentially attend to the eyes of racial ingroup members (i.e., White faces) as

opposed to racial outgroup members (i.e., Black faces), (b) whether White participants preferentially attend to the eyes of minimal ingroup members as opposed to minimal outgroup members, in the absence of any differences in facial features and pre-existing perceptual experience, and (c) whether inducing individuation motives in White participants increases their attention to the eyes of racial outgroup members (i.e., Black faces). While the originally published findings responded with "yes" to all three research questions, subsequent re-analysis of Kawakami et al.'s (2014) results revealed that only the first "yes" is warranted after controlling for different base rates in attention allocation (i.e., the fact that participants could have looked longer at one face over the other), which is important as ingroup and outgroup faces were simultaneously presented on the screen (Correll & Hudson, 2020). In this re-analysis, it was confirmed that White participants attended more to the eyes of White faces than Black faces (see Goldinger et al., 2009 for similar evidence in a different interracial context; see Friesen et al., 2019 for an extension of this work to the domain of smile judgments). However, White participants did not show a different pattern of preferential attention to any specific facial feature of minimal ingroup members' faces to that shown for minimal outgroup members' faces, but only an overall visual attention preference for minimal ingroup faces over minimal outgroup faces. Finally, White participants' attention to the eyes of Black faces did not increase following the manipulation of individuation motives, even though their own-race bias was attenuated. Furthermore, while not dealing with group membership *per se*, another set of experiments from Kawakami and colleagues (2021) is important to mention. Specifically, in two experiments, they found that participants looked longer at the eyes of those with more similar scores to them on a personality test, regardless of which racial group those others belonged to (e.g., Black or White faces). The authors interpreted this finding as an example of how perceivers' motivation can influence person perception in a top-down manner. This finding is particularly important given that

minimal groups are frequently created based on personality tests, implying participants' greater similarity to minimal ingroup versus minimal outgroup members. In summary, the socio-cognitive account emphasises that at baseline, people have stronger motivation to process individuating features of own-race faces than other-race faces, but that these individuation motives can be increased towards other-race faces as well, resulting in reduced own-race bias. However, the exact mechanisms through which individuation motives lead to attenuation of own-race bias are currently not well understood.

As social psychological research has almost exclusively examined racial groups' influence on visual attention to faces, this introduction has focused on race to offer examples of how attentional shifts to faces and specific facial features can be studied in intergroup contexts. To the best of my knowledge, no work other than the work of Kawakami and colleagues (2014) has examined how *minimal* group membership influences visual attention to faces, and no previous work has examined minimal group membership's influence on visual attention to faces in the context of emotion judgments. Thus, given that the findings concerning racial groups cannot be generalised to minimal groups and that no minimal group research is available to directly build on, the hypotheses listed below have mostly been generated on the basis of research examining visual attention to faces and facial expressions of emotion in the absence of any group distinctions in combination with broader theoretical assumptions about how face processing might shift once cues of minimal group membership and coalitional relations are introduced. For this reason, all hypotheses should be treated as exploratory.

Hypotheses. Holistic processing of faces, wherein facial features are integrated into a compact facial template, is often claimed to be the default mode of face processing (Richler et al., 2009; Richler et al., 2011). However, social psychological literature has provided some evidence that holistic processing of faces might vary with both racial and minimal group

membership. Specifically, multiple experiments suggested that ingroup members' faces are processed more holistically than those of outgroup members, but these conclusions were reached only on the basis of self-report paradigms (e.g., inversion and composite face paradigms, Hugenberg & Corneille, 2009; Michel et al., 2007), while the same mechanistic explanation has never been explored in the domain of visual attention to faces. This means that if one aims to search for signs of reduced holistic processing of outgroup relative to ingroup faces in the domain of visual attention to faces, one first needs to look at vision science literature to see if any indices of holistic face processing have been proposed to date. Having done that, one will find that holistic processing of faces has most often been indexed by fixations to the centre of the face (i.e., nose), which allows one to peripherally grasp information from the whole face (Blais et al., 2008; Hsiao & Cottrell, 2008), and by a higher number of inter-featural saccades, which reflects exploration of multiple facial features as opposed to in-depth processing of a restricted section of the face (Bombari et al., 2009; Schmid et al., 2011). The first hypothesis connects these two completely independent strands of research.

H1. If holistic processing is indeed disrupted for outgroup relative to ingroup faces and if holistic processing is indeed captured by fixations to the nose and a greater number of inter-featural saccades, then one could expect to observe longer fixations and more viewing time to the nose of ingroup members in comparison to outgroup members and to observe more inter-featural saccades when exploring ingroup members as opposed to outgroup members' facial expressions. Importantly, even though this hypothesis centres around group category membership to maintain connection with previous literature, coalitional relations might similarly change visual attention patterns.

Another possibility builds on the literature which emphasised that observers tend to look at facial features diagnostic of the emotion that is being expressed. Specifically, mouth

grabs more attention in prototypical expressions of happiness, while eyes grab more attention in prototypical expressions of anger (Bombari et al., 2013; Eisenbarth & Alpers, 2011; Scheller et al., 2012; Schurgin et al., 2014; Smith et al., 2005). This pattern of visual attention to faces is particularly likely in the current experiment not only because Experiment 3 relies on static and prototypical facial stimuli akin to those in the cited literature, but also because facial stimuli reflecting anger have been specifically preselected to have higher action unit activation in the eye region (i.e., brow lowerer) while facial stimuli reflecting happiness have been preselected to show an overt smile, and therefore have higher action unit activation in the mouth region (i.e., lip corner puller). This pre-selection procedure makes eye and mouth regions particularly salient in the angry and happy facial expressions respectively. Furthermore, based on the results from Experiments 1 and 2, one can hypothesise that ingroup positivity bias and cooperation positivity bias in people's judgments of facial expressions could be preceded by a corresponding *attentional search* for signs of positive emotions in ingroup members and those with whom one is to cooperate. The second hypothesis captures this possibility.

H2. If participants engage in an *attentional search* for signs of positive emotions in ingroup members and those with whom one is to cooperate, one could expect to see longer individual fixations and dwell time to the eye region of outgroup members and competitive others relative to ingroup members and cooperative others when they are expressing anger, but longer individual fixations and dwell time to the mouth region of ingroup members and cooperative others relative to outgroup members and competitive others when they are expressing happiness.

In the introduction of Experiment 3, I also described findings suggesting that attention to the eyes increases with increased similarity and shared racial group membership (Kawakami et al., 2014; Kawakami et al., 2021). However, this work has only looked at how

attention is allocated to faces when they are not expressing any emotions, making it difficult to predict how this preference for the eyes of similar others might change when different emotional expressions are introduced, particularly those that signal interpersonal threat (i.e., anger, Ewbank et al., 2009).

One final possibility was originally discussed by Lazerus and colleagues (2016) who believed that "an account under which group membership is somehow distorting perception is far less likely than the account under which group membership is biasing judgments of the semantic valence representations" (p. 1124). Specifically, the authors interpreted ingroup positivity bias as a consequence of a spillover of positivity from the ingroup to the domain of judgments of facial expressions (see also Dunham, 2018), rather than a consequence of any particular visual attention mechanism. This interpretation is indistinguishable from the classical idea of *affective priming* (Fazio, 2001; Houwer, 2009), where the evaluative response to expectations of shared group category membership with the expresser translates to corresponding valence judgments of facial expressions. The third hypothesis captures this possibility.

H3. Minimal group membership and coalitional relations will not affect visual attention to faces at all but will instead exert their influence only on post-perceptual judgments of otherwise similarly explored facial expressions.

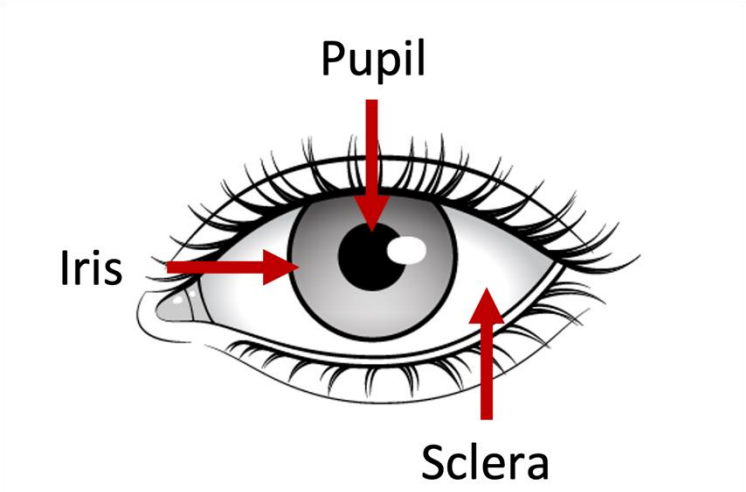
Peripheral physiology and arousal

In Experiment 3, I measured two physiological responses known to index arousal: (a) changes in pupil size, and (b) skin conductance response. In two separate sections dedicated to each of these responses, I will briefly introduce the neural pathways behind their occurrence and then proceed to discuss research which has relied on these responses to provide insights in the domain of interpersonal and intergroup emotion processing.

Pupil size. The pupil is the black round part at the centre of the iris (see Figure 2.8) and all changes in its size are controlled by two largely distinct neural pathways, the *parasympathetic pathway*, which is responsible for pupil constriction, and the *sympathetic pathway*, which is responsible for pupil dilation (Mathôt, 2018). On the most peripheral ends of these pathways, there are two muscles called *sphincter pupillae* and *dilator pupillae*, innervated by the parasympathetic and the sympathetic nervous system respectively. On one side, sphincter pupillae, located immediately around the pupil, leads to its constriction by contracting and thus "squeezing" the pupil. On the other side, dilator pupillae, located on the outskirts of the iris, leads to pupil dilation by contracting and thus pulling the iris outward and "stretching" the pupil. Given that pupil dilation is a sympathetically driven response, it has been shown to correlate with other indices of sympathetic neural system activity (e.g., skin conductance response, Bradley et al., 2008; Schumann et al., 2020; Wang et al., 2018). More details of the two neural pathways can be found in Beatty and Lucero-Wagoner (2000), Lykstad et al. (2022), and Mathôt (2018).

Figure 2.8

Basic anatomy of the eye



The pupil constricts in response to brightness and dilates in response to darkness (i.e., pupillary light reflex), constricts when people fixate on a near-by object and dilates when people fixate on a distant object (i.e., pupil near response) and finally and most importantly for Experiment 3, dilates in response to increased physiological arousal or mental effort (Mathôt, 2018). While the pupil light response and the pupil near response are examples of how pupil size is affected by low-level features of the sensory environment (i.e., brightness and proximity respectively), the third category of pupil size changes also reflects the input of other, more "psychological" stimuli. For example, when brightness and proximity are methodologically or statistically controlled, the pupil has been shown to dilate when people's working memory load is increased (Hess & Polt, 1964; Peysakhovich et al., 2015; Mathôt et al., 2015), leading to a consensus that pupil dilation can be interpreted as an indicator of increased mental effort. Furthermore, pupil dilation is considered to be an indicator of increased physiological arousal given that it occurs when people view arousing images or listen to arousing sounds regardless of their valence (Bradley et al., 2008; Hess & Polt, 1960; Hess et al., 1965; Partala & Surakka, 2003) and when people's expectations are violated (Brod et al., 2022; Hepach & Westermann, 2013; Hepach et al., 2016; Kloosterman et al., 2015; Krüger et al., 2020; Lavin et al., 2014; Preuschoff et al., 2011). Two important points about effort- and arousal-fuelled pupil dilation should be noted. First, effort- and arousal-fuelled pupil dilation is usually slower in comparison to other categories of pupil size changes (i.e., pupillary light reflex and pupil near response, Mathôt, 2018). Second, effort- and arousal-fuelled pupil dilation tends to be much smaller in magnitude in comparison to the other categories of pupil size changes and tends to display much more individual variability (Cuve, Stojanov, et al., 2021). Finally, even though this "slower" category of pupil size changes is most often explicitly interpreted in terms of mental effort or physiological arousal,

a broader rule of thumb is that "whatever activates the mind causes the pupil to dilate" (Mathôt, 2018, p. 13).

To date, pupillometry (the study of changes in pupil size, Sirois & Brisson, 2014), remains to a considerable extent under-used in the fields of interpersonal and, especially, intergroup emotion processing. Nevertheless, a few findings from these fields emphasise the intrinsic importance of facial expressions of emotion and narrow down the scope of potential hypotheses in Experiment 3. For instance, both adults' (Aktar et al., 2021; see also Partala & Surakka, 2003) and infants' (Prunty et al., 2022) pupils have been shown to dilate more in response to facial expressions of emotion than in response to neutral facial expressions. However, research examining whether certain facial expressions of emotion trigger stronger pupil dilation than others has yielded somewhat inconsistent results, with some studies indicating the strongest pupil dilation after angry facial expressions and interpreting these as a response to interpersonal threat (Kret et al., 2013) and other studies indicating the strongest pupil dilation after happy facial expressions and interpreting these as a response to interpersonal reward (Jessen et al., 2016, see also Prunty et al., 2022; Yrttiaho et al., 2017). The findings of Hepach and Westermann (2013) have extended this research in an important direction by showing that infants' pupils not only dilate more in response to one emotional expression (i.e., anger) over the other (i.e., happiness), but also in response to facial expression-action incongruence. Specifically, 14-month-old infants' pupils dilated more when they observed an angry person performing a positive action or a happy person performing a negative action than they did for congruent emotion-action combinations, suggesting that infants are able to associate basic emotions with typical actions and are responsive when these associations are violated.

The pupillometric research discussed so far has documented adults' and infants' general responsivity to facial expressions of emotion as well as their sensitivity to

expectancy-violating stimuli (see also Proulx et al., 2017), but to the best of my knowledge, no research to date has examined how pupil size changes when people are presented with real-world or minimal ingroup and outgroup members expressing different facial expressions of emotion. However, it is not difficult to imagine that people harbour emotion-group (and emotion-coalition) associations similar to the emotion-action associations documented by Hepach and Westermann (2013) and are similarly responsive to violations of these associations (Freeman et al., 2018). In this regard, as discussed in the introduction of this chapter, minimal group membership has indeed been shown to influence early visual processing of facial expressions, with expressions which violate prior expectations (ingroup anger and outgroup joy) eliciting larger EEG activity 30 to 100 ms after onset than expressions which do not violate prior expectations (ingroup joy and outgroup anger, Gamond et al., 2017). Given that the pupil has been shown to dilate when people's social expectations are violated, my hypotheses in the domain of physiological responses revolve around multiple potential ways in which different pieces of social information – group category membership, coalitional relations, emotion category – could be perceived as incongruent and therefore lead to increased physiological arousal in Experiment 3. Finally, as both pupil dilation and skin conductance response have a sympathetic neural basis, I have the same hypotheses for both physiological responses and these hypotheses will be listed at the end of the next section. However, it should be noted that changes in pupil diameter over the course of a trial in an experiment usually reflect both pupil constriction, which is controlled by the parasympathetic nervous system, and pupil dilation, which is controlled by the sympathetic nervous system, while skin conductance response is only controlled by the sympathetic nervous system. Thus, while I do not have a reason to specify different hypotheses for each physiological response, reports of different patterns of effects on pupil

dilation and skin conductance response are not unusual in the literature (see Leuchs et al., 2019; Wang et al., 2018 for examples).

Skin conductance response. Skin conductance, also known as electrodermal activity, refers to changes in electrical properties of the skin due to sweat secretion (Boucsein, 2012). Specifically, activation of the sympathetic nervous system leads to increased activity of the sweat glands which increases the electrical conductance of the skin because sweat mainly consists of water and electrolytes (Hess, 2011). Skin conductance dynamics are usually split into tonic and phasic components. The tonic component (i.e., skin conductance level, SCL) refers to variations in skin conductance that occur slowly and are observed in the absence of any known triggers (Benedek & Kaernbach, 2010a). SCL is most often related to thermoregulation and general arousal in the literature (Amin & Faghih, 2019). The phasic component (i.e., skin conductance response, SCR) refers to a temporary increase in skin conductance which occurs within a few seconds following an external – novel or otherwise activating – stimulus and is characterised by a steep initial incline followed by a slower return to baseline (Kuhn et al., 2022). Psychological research primarily examines SCR in the context of known external stimuli (i.e., event-related SCR), interpreting its occurrence and amplitude as indices of varying sympathetic arousal in response to presented stimuli, but SCR can also occur spontaneously (i.e., non-specific SCR, Hess, 2011).

Thus far, increased SCR has been documented in a variety of situations informative for hypothesis development in Experiment 3. First, increased SCR has most consistently been observed in response to angry facial expressions (Clark et al., 1992; Folz et al., 2022; Nava et al., 2016). For instance, a recent study by Folz and colleagues (2022) documented an increase in skin conductance after participants were exposed to angry in comparison to neutral, happy, sad, and fearful facial expressions with the early peak of this response suggesting faster processing of anger in comparison to other emotions, potentially due to its threat-related

implications. Stronger SCR in response to angry as opposed to happy facial expressions have also been identified in infants as young as 3-4 months of age (Nava et al., 2016). Even though threat has always been at the core of interpretations offered to explain increased SCR to angry facial expressions, other research has more explicitly related increased SCR to feelings of threat. Specifically, Tsikandilakis and colleagues (2020) showed that increased SCR specifically followed facial expressions of hostility, wherein participants inferred that the expresser had the intent to inflict physical harm, as opposed to facial expressions of anger, wherein participants detected signs of frustration and stress in the expresser. Furthermore, SCR has been shown to gradually increase as one is approached by an unfamiliar person (Candini et al., 2021). Together, these results more unambiguously suggest that cues of interpersonal threat, such as an unknown person's proximity or facial expressions of anger and hostility, are particularly likely to trigger increased sympathetic arousal.

Second, SCR has been shown to index arousal responses to unexpected and surprising outcomes (i.e., prediction errors), similar to pupil dilation, but the empirical evidence for this remains mostly restricted to the domain of aversive learning (Stemerding et al., 2022; Willems & Vervliet, 2021). For instance, in one aversive learning paradigm, stronger SCR occurred on trials where the unconditioned stimulus (e.g., brief electric shock) was unexpectedly presented or unexpectedly omitted relative to trials where the presentation or omission of the unconditioned stimulus was expected (Stemerding et al., 2022). Finally, given that SCR has not thus far been used to index physiological arousal during intergroup emotion processing or intergroup interactions more broadly (but see Brown et al., 2006), hypotheses listed below mostly rely on presented findings documenting SCR's responsivity to violations of expectations.

Hypotheses. SCR and changes in pupil size were used to index implicit processing of social information about the expresser in two sections of each trial – (a) during the

presentation of the textual cue, which introduced group category membership of, and coalitional relations with, the upcoming expresser in the absence of facial or any other cues, and (b) during the subsequent presentation of the target's neutral and emotional facial expression. As the textual cue incorporates only two factors of interest – group category membership and coalitional relations – while the facial expressions of emotion incorporate all three factors of interest – group category membership, coalitional relations, and emotion category – with emotion category acutely salient, these two sections of each trial warrant separate hypotheses.

First, if SCR or changes in pupil size are observed after the onset of the textual cue, they can only be interpreted as indexing increased physiological arousal in response to concepts of "ingroup members", "outgroup members", "cooperation", "competition" and their congruence or incongruence. As such, I hypothesised that textual cues would be the section in which any responses stemming from group-coalition incongruence are particularly likely to arise. Specifically, based on the coalitional preferences reported in Experiments 1 and 2 as well as previous research (Balliet et al., 2014; Brewer, 2008; Yamagishi & Kiyonari, 2000), I hypothesised that participants would be more likely to expect cooperative within-group interactions and competitive between-group interactions. These expectations are potentially amplified by the characteristics of the experimental design: (a) by the usage of the word "teams" instead of "groups" (e.g., team Blue and team Red as opposed to group Blue and group Red) as well as (b) the general gamified environment of Experiment 3. This reasoning is captured by the first hypothesis.

H1. If participants indeed harbour ingroup-cooperation and outgroup-competition associations, one could expect to observe a two-way group category membership x coalitional relations interaction such that participants show increased pupil dilation and SCR

in response to ingroup members with whom they are to compete and outgroup members with whom they are to cooperate in the future.

Second, hypotheses pertaining to changes in physiological arousal during the presentation of facial expressions of emotion are more exploratory due to the complex interplay between different pieces of social information offered to participants before and during this trial section. Focusing on the interplay between targets' group category membership, coalitional relations and expressed emotion, I hypothesised that participants' physiological responses would reflect one of the following:

H2. a two-way interaction between group category membership and emotion category, such that participants display greater physiological arousal in response to incongruent group-emotion combinations (ingroup-anger, outgroup-happiness) as opposed to congruent ones (ingroup-happiness, outgroup-anger); or

H3. a two-way interaction between coalitional relations and emotion category, such that participants display greater physiological arousal in response to incongruent coalition-emotion combinations (expected cooperation-anger, expected competition-happiness) as opposed to congruent ones (expected cooperation-happiness, expected competition-anger).

In other words, given that both ingroups and cooperation carry more positive connotations relative to outgroups and competition, both of these factors could theoretically be congruent or incongruent with positive and negative emotion categories. The second and third hypotheses focus on group-emotion and coalition-emotion incongruence respectively, rather than on group-coalition incongruence, because emotion category is the only piece of social information that was acutely salient upon presentation of the facial expression. In summary, based on previous literature, I hypothesised that changes in pupil size and SCR might reveal an implicit processing step (e.g., participants' responsivity to social

incongruence) that was not evident in participants' judgments of facial expressions in Experiments 1 and 2 (see Gamond et al., 2017).

Method

Participants. Participants were 36 adults (58.3% women, 41.7% men) between the ages of 18 and 60 (M_{AGE} = 29.5, SD_{AGE} = 11.63), recruited using the university participant network in February 2020 (N = 9) and September 2022 (N = 27). These lab-based participants' valence, arousal, and genuineness judgments of facial expressions have already been analysed as part of a bigger sample in Experiment 1. Thus, Experiment 3 instead focuses on participants' gaze patterns and physiological responses (e.g., changes in pupil size, SCR) as real-time indicators of participants' implicit processing of group- and coalition-related information as well as facial expressions of emotion. One participant's SCR was not analysed as they were found to be repeatedly peeling the sensors off their fingers during the emotion task which resulted in an excessive number of artifacts. Therefore, the final sample whose SCR was analysed included 35 participants, while the final sample whose gaze patterns and pupil size were analysed included all 36 participants.

Procedure. The procedure of Experiment 3 corresponds to the lab-based procedure of Experiment 1. To recap, prior to coming to the lab, participants completed a pre-lab-visit questionnaire which assigned them to minimal groups, introduced them to the two versions of the shooting game, and included the TAS-20 (Parker et al., 2003), the DASS-21 (Lovibond & Lovibond, 1995), questions about participants' own intra- and intergroup coalitional preferences as well as demographic questions and questions about participants' eyes and vision. Upon their arrival to the lab, participants read the instructions which reminded them of their group category membership and available shooting game versions, after which they completed the emotion task while their gaze and physiological responses were recorded. Having completed the emotion task, participants filled in an exit questionnaire probing their

assumptions about the research questions and the cover story, after which they were debriefed. The pre-lab-visit questionnaire, reminder instructions and exit questionnaire were hosted on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), while the emotion task was displayed in PsychoPy (version 3.2.4; Peirce, 2007).

Apparatus, pre-processing, and analysis. The lab-based emotion task was run on a Dell computer (Intel Core i7-3610QM @ 2.30 GHz, 16 GB RAM, Windows 10). The facial stimuli (539 x 768 pixels) were presented a 53 x 30 cm monitor (60 Hz refresh rate, 1920 x 1080 pixels, 45.99 x 27.01 degrees of visual angle). The emotion task was completed in a dimly lit, sound-proof testing room. Below, I describe the devices used to record gaze coordinates, pupil size, and SCR, followed by pre-processing steps, and statistical analyses.

Gaze coordinates and pupil size. Gaze coordinates and pupil size were continuously recorded using the Gazepoint GP3-HD eye-tracker, recording at 150 Hz, controlled through a custom Python script in PsychoPy (Peirce, 2007). The eye-tracker was placed at a 45° angle and 60-65 cm from the participants' eyes ($M_{DISTANCE}$ = 63.19 cm) in line with the Gazepoint manual. The Gazepoint Control software, and physical measurement were used to aid set-up and find an optimal position for each participant. Participants completed the emotion task while resting their head on a chinrest, used to minimise head movements and to standardise the distance from the participants' head to the monitor and the eye-tracker. Participants were also instructed to avoid sudden head and body movements, although they could move their eyes freely to explore the facial stimuli.

Prior to starting the emotion task, a nine-point calibration was administered. Calibration criteria for continuing with the task were determined *a priori*: (a) all nine calibration points had to be deemed valid by the Gazepoint Control software; and (b) average calibration error had to be below or equal to 1 degree of visual angle (~ 40 pixels) as recorded by Gazepoint Control software. While the first criterion was met by all calibrations, if for

whatever reason, a sufficiently low average calibration error could not be obtained after five calibration attempts, the experiment continued anyway. Given the presence of participants for whom a sufficiently low average calibration error could not be obtained for all calibrations and given that the displacement of fixation registrations tends to increase over time (Carr et al., 2022), I applied a drift correction. This means that the position of individual gaze samples was corrected for small drifts in the calculation of gaze position (Mathôt et al., 2014). Specifically, I calculated the vertical and horizontal distances between a known target in space presented before each facial expression (i.e., the fixation cross) and the gaze position estimated by the eye-tracker used these distances to correct the position of individual gaze samples. Most importantly, the results did not change regardless of the whether fixation positions were corrected or not. Finally, to account for the deterioration of calibration quality over time (Holmqvist et al., 2012), recalibration was performed after every 32 trials[32] (~ 20 minutes), and additionally whenever deemed necessary (e.g., when participants closed their eyes for an unusually long period or when they had moved their head off the chinrest).

Pre-processing and analysis of gaze and pupil data was done in R (version 4.1.3). Textual cues (N = 356) and facial stimuli (N = 137) where more than 40% of gaze samples were lost were excluded. Additionally, facial stimuli on trials where participants did not accurately respond to the two preceding attention checks were not analysed[33]. In order to analyse gaze patterns, raw eye-tracking data must be parsed into fixations and saccades. This was achieved using Gazepath, a data-driven velocity-based algorithm that accounts for individual differences and differences in data quality (Geller et al., 2020). Given that analysis of saccade properties is affected by the low sampling rate of the GP3-HD eye-tracker (Cuve,

[32] For lab-based participants who completed the experiment in 2020, recalibrations were performed after every 36 trials.

[33] The same exclusion criterion was not applied to textual cues because textual cues contained information participants later needed to report in attention checks.

Stojanov et al., 2021) and that most social psychological research which informed my proposed hypotheses has only conducted fixation analyses, Experiment 3 focused primarily on fixation metrics and only calculated the percentage of intra- and inter-featural saccades on the basis of location of each of two adjacent fixations within the same trial (e.g., if two adjacent fixations landed on the same face region, one can assume that an intra-featural saccade took place between them, and if two adjacent fixations landed on different face regions, one can assume that an inter-featural saccade took place between them).

After fixations were identified, they were assigned to so-called areas of interest (AOIs). Following the most common practice of eye-tracking research in the domain of faces, three AOIs were pre-defined in Experiment 3: eyes, nose, and mouth. Individual fixations were then assigned to these AOIs using the limited-radius Voronoi tessellation method, which has been shown to be the most robust method of defining AOIs, particularly in the domain of facial stimuli (Hessels et al., 2016). As applied in the current experiment, the limited-radius Voronoi tessellation method can be divided into four steps. Under the first step, facial landmarks of all stimuli were identified using OpenFace (Baltrusaitis et al., 2018) and centres of the four AOIs – left eye, right eye, nose, and mouth[34] – were determined on the basis of these facial landmarks (see Figure 2.9A for an example of the OpenFace facial landmark map being applied to one emotion task stimulus and Figure 2.9B for the centres of each AOI, indicated by the red dots). Under the second step, Euclidian distances between gaze sample coordinates and centres of AOIs were calculated and gaze samples were assigned to the AOI with the smallest Euclidian distance. Regardless of whether the centre of the left or the right eye had the smallest Euclidian distance to the gaze sample, it should be

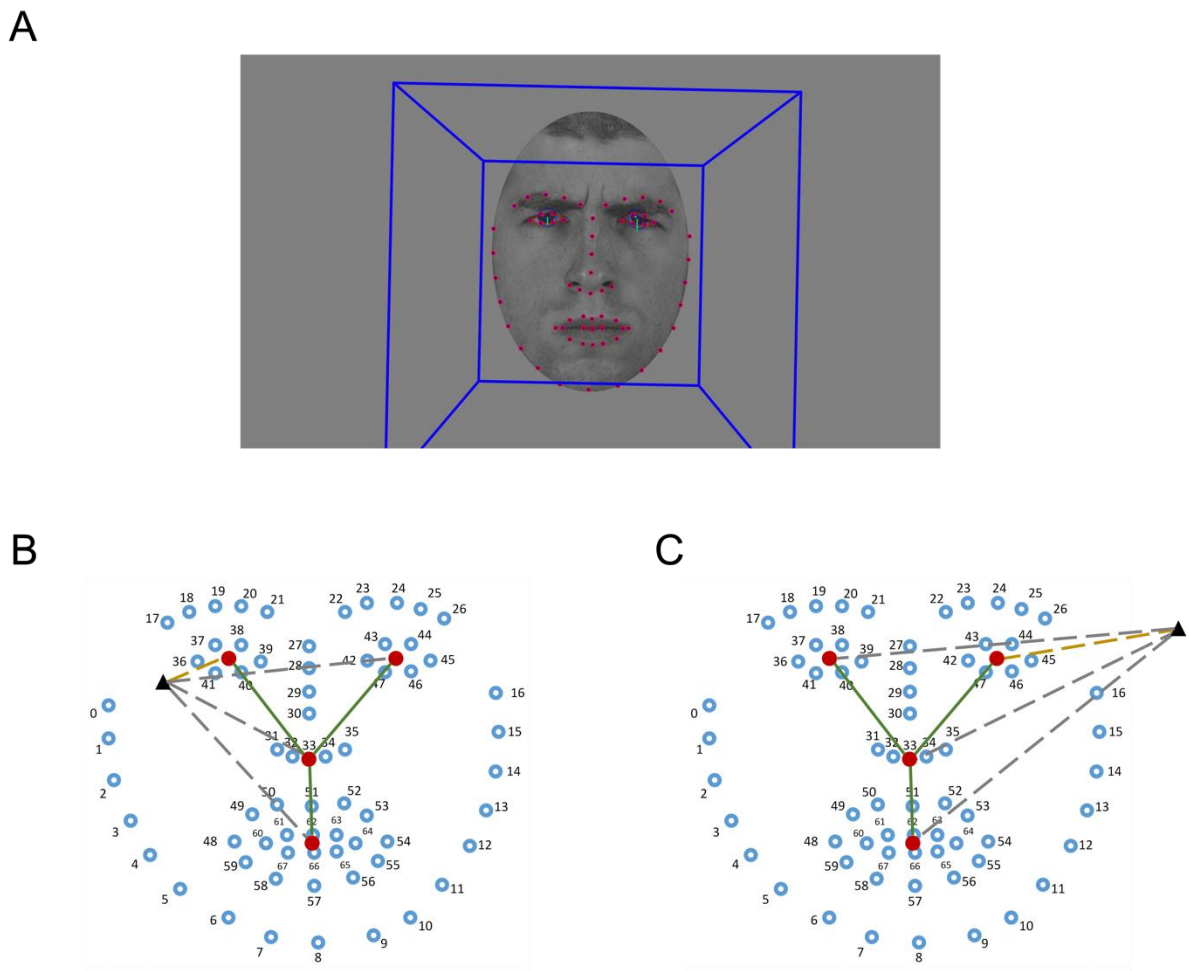
[34] Left eye and right eye were only separated in this first step as it was necessary to track the distance of participants' gaze samples from the centres of both eyes. Regardless of whether the individual gaze sample was closer to the left or the right eye, it was always assigned to the broader AOI "eyes".

noted that the gaze sample was always assigned to the AOI "eyes". Under the third step, if the smallest Euclidian distance between the gaze sample and the centre of a single AOI was larger than the average Euclidian distance between adjacent AOIs, the gaze sample was not assigned to an AOI and was subsequently excluded from analysis (see Figure 2.9)[35]. Under the fourth step, if the Euclidian distances between an individual gaze sample and centres of two or more AOIs were roughly the same (i.e., difference score <= 10 pixels, which corresponds to 0.25 degrees of visual angle), suggesting that the gaze sample is exactly on the border between two or more AOIs, this gaze sample was also excluded.

[35] Size of an AOI is a subjective choice, and a single empirically tested rule of thumb a researcher can abide by suggests erring towards defining larger AOIs (Hessels et al., 2016), which is what the approach described in the main text attempts to achieve. However, to be sure that the way I restricted AOI radii does not affect the results, I also adopted two other approaches, always relying on inter-AOI Euclidian distances as a guiding framework to avoid imposing an arbitrary cut-off. Under the first approach, I excluded gaze samples whose smallest Euclidian distance to an AOI was larger than the distance between this AOI of interest and an adjacent AOIs (e.g., the distance between eyes and nose if the gaze sample was assigned to the AOI "eyes", or the distance between nose and mouth if the gaze sample was assigned to the AOI "mouth"). Given that the distance between the tip of the nose and the centre of the mouth is usually smaller than the distance between the tip of the nose and the centre of the eye, this approach results in unequally sized AOIs that are better tuned to the distances between facial features. Under the second approach, I excluded gaze samples whose smallest Euclidian distance the centre of a single AOI was still larger than the smallest observed Euclidian distance between the centres of all different AOIs. This approach is the most stringent, resulting in higher confidence that the analysed gaze samples indeed fall onto narrowly defined AOIs. However, this approach also results in the highest data loss (i.e., highest number of gaze samples excluded from the analysis), potentially reducing statistical power to detect meaningful effects. Additionally, under the latter approach, gaze samples originally assigned to the eye region are disproportionately penalised and for this reason, any minor differences in results after adopting this approach and the other two approaches, mostly pertaining to how gaze was allocated between targets' eyes and nose, will not be further discussed. Most importantly, results pertaining to targets' group category membership and coalitional relations did not change regardless of how AOI radii were defined.

Figure 2.9

Assignment of fixations to areas of interest



Notes. (A) Facial landmark map of OpenFace (Baltrusaitis et al., 2018) applied to an example emotion task stimulus. (B & C) OpenFace facial landmark map. Centres of AOIs – left eye, right eye, nose, mouth – are indicated by red dots. Two hypothetical gaze samples are indicated by two black triangles while Euclidian distances from the respective gaze samples to each AOI are indicated by dashed lines. The differently coloured (ochre) dashed line indicates the AOI to which each gaze sample was initially assigned. (B) Since the Euclidian distance between the gaze sample and the targets' right eye is smaller than the average Euclidian distance between AOIs (i.e., average length of green lines), this gaze sample remains assigned to the AOI "eyes". (C) Since the Euclidian distance between the gaze sample and the targets' right eye is larger than the average Euclidian distance between AOIs, this gaze sample is excluded from analysis.

Pupil data were first cleaned by removing pupil sizes that were more than three standard deviations away from the average pupil size for each participant. This threshold was determined by visualising both the raw and the smoothed pupil size time-series. Missing data (e.g., due to blinks) were extended by 150 ms and were imputed using a linear interpolation

implemented in R package GazeR (Geller et al., 2020). After interpolation, I dealt with any unrealistically rapid pupil size changes by calculating normalised dilation speed (i.e., absolute change in pupil size between adjacent samples divided by the time between them), flagging and ultimately removing speed values above the specified sensitivity threshold (i.e., median absolute deviation * 16[36]). A subtractive baseline correction was then applied (Mathôt et al., 2018). Specifically, median pupil size during the last 20 samples of the preceding trial section (e.g., fixation cross) and the first 20 samples of the current trial section (approximately 260 ms) was taken as baseline pupil size, from which all individual pupil sizes were subtracted on each trial. Finally, as luminance is the main determinant of pupil size, it is important to emphasise that all facial images were equalised for luminance using the SHINE toolbox in MATLAB (Willenbockel et al., 2010).

Skin conductance. Recordings of skin conductance were obtained from the BIOPAC M160 system with the Electrodermal Activity BioNomadix module, sampling at 2000 Hz. The disposable isotonic gel electrodes were attached to the distal phalanges of participants' index and middle fingers on their non-dominant hand. The skin where the SCR electrodes were placed was cleaned with water and dried with a cotton tissue. Before starting the emotion task, participants were asked to take a deep breath and hold it for three seconds as this frequently raises the skin conductance response (i.e., phasic component of skin conductance) to the level observable by the human eye (Figner & Murphy, 2011; Greco & Baenninger, 1991). Only two participants did not initially show the expected skin conductance response after taking a deep breath, but in both of these cases the expected response appeared following the change of electrodes.

[36] The constant controls the sensitivity threshold – the higher the value, the more extreme the normalised pupil speed needs to be to be flagged for removal. Authors of GazeR (Geller et al., 2020), where these calculations are implemented, recommended 16 as the value of the constant and set it as the default.

For analysis, skin conductance was first downsampled to 10 Hz using the Acqknowledge software (Braithwaite et al., 2013). A continuous decomposition analysis (CDA), which decomposes skin conductance into tonic activity (i.e., the slowly changing component of skin conductance level, SCL) and phasic activity (i.e., the quickly changing component of skin conductance response, SCR) was then conducted in Ledalab (version 3.4.9, Benedek & Kaernbach, 2010), running in Matlab 2020a. As part of the Ledalab pre-processing steps, I applied adaptive smoothing, specified the response window as 1-6s following stimulus onset and defined the minimum threshold of SCR at 0.01 μS.

Analysis strategy. Participants' gaze patterns were analysed only during the periods when they were presented with the neutral and emotional facial expression stimuli[37]. Once fixations had been assigned to AOIs (eyes, nose, or mouth), two fixation metrics were calculated for each AOI on a trial-by-trial basis: (a) average fixation duration (i.e., how long individual fixations on a particular AOI lasted on average), and (b) dwell time (i.e., how long participants looked at a particular AOI in total). Increases in both average fixation duration and dwell time signal greater visual attention to a particular AOI. Dwell time was transformed to proportions (e.g., dwell time on eyes / total dwell time on all AOIs) on a trial-by-trial basis prior to being included in the analyses.

Two mixed models tested the fixed effects of targets' group category membership (levels: ingroup member, outgroup member), coalitional relations (levels: competitive, cooperative), emotion category (levels: angry, ambiguous, happy) and AOI (levels: eyes, nose, mouth) on average fixation duration (linear mixed model) and dwell time proportions (generalised mixed model with a beta distribution and a logit link function). The linear mixed

[37] Neutral facial expressions were always presented for 500 ms before emotional facial expressions to simulate facial movement corresponding to initial expression development and were never intended to be analysed separately.

model with average fixation duration as the dependent variable included random intercepts for participants and targets' identities as well as a correlated random slope for AOI within participant, while the generalised mixed model with dwell time proportions as the dependent variable only included by-participant and by-target-identity random intercepts.

Finally, I calculated the number of intra-featural saccades, inter-featural saccades, and "out-of-AOI" saccades. A saccade was labelled as intra-featural if two adjacent fixations fell on the same AOI (eyes, nose, or mouth), while a saccade was labelled as inter-featural if two adjacent fixations fell on different AOIs (e.g., fixation 1 fell on "eyes" and fixation 2 fell on "nose"). If a saccade was between two fixations where at least one of them was not assigned an AOI, indicating that the fixation was not close enough to any of the three AOIs to be assigned to them, this saccade was coded as "out-of-AOI". These saccade numbers were then transformed to proportions (e.g., number of inter-featural saccades / total number of saccades) on a trial-by-trial basis prior to inclusion in the analyses. The final mixed model (generalised mixed model with a beta distribution and a logit link function) tested the fixed effects of targets' group category membership (levels: ingroup member, outgroup member), coalitional relations (levels: competitive, cooperative), emotion category (levels: angry, ambiguous, happy), and saccade type (levels: intra-featural, inter-featural, out-of-AOI) on saccade proportions. This model also included by-participant and by-target-identity random intercepts. The results remained unchanged even when out-of-AOI saccades were excluded from proportion calculations and subsequent models.

To reiterate, changes in participants' pupil size were analysed in two sections of each trial – during the presentation of the textual cue (5000 ms), and during the presentation of the neutral (500 ms) and the emotional facial expression (5000 ms). Following the approach of Mathôt et al. (2017), each 13.34-ms window (equal to two adjacent gaze samples) of baseline corrected pupil size recordings was analysed using linear mixed models and a significance

threshold of at least 200 contiguous milliseconds in which $p < .05$ was used. This means that the effect needed to meet the $p < .05$ criterion in at least 15 linear mixed models on adjacent 13.34-ms windows to be considered statistically significant, and not a Type I error (i.e., a false positive). The significance threshold of at least 200 contiguous milliseconds is used instead of traditional corrections for multiple comparisons (e.g., Bonferroni correction) which assume that observations are uncorrelated, because this is not the case with pupil size data (Mathôt & Vilotijević, 2022). In the analysis of pupil responses during the presentation of the textual cue, the linear mixed model that was specified in each 13.34-ms window included targets' group category membership (levels: ingroup member, outgroup member) and coalitional relations (levels: competitive, cooperative) as fixed effects and by-participant intercepts as a random effect. In the analysis of pupil responses during the presentation of the facial expressions, the linear mixed model specified in each 13.34-ms window included targets' group category membership (levels: ingroup member, outgroup member), coalitional relations (levels: competitive, cooperative) and emotion category (levels: angry, ambiguous, happy) as fixed effects and by-participant intercepts as a random effect[38]. In addition to all facial stimuli being equalised for luminance and spatial frequency, I took another precautionary measure to make sure that any detected pupil changes do not stem from luminance confounds. Specifically, I calculated both (a) average luminance of each textual cue and each facial stimulus and (b) local luminance of the mouth region (490 x 217 pixels) of each facial stimulus as the visibility of teeth was the only potential source of luminance differences between the facial stimuli. I then controlled for these variables in separate linear

[38] By-trial random intercepts were not included given the subtractive baseline correction (Mathôt et al., 2018), and the random slopes of trial on participants were not included across measures of peripheral physiology for the same reason as in Experiment 1.

mixed models. Results remained the same regardless of whether and how luminance was controlled for.

Similar to changes in pupil size, SCR was identified and analysed in two sections of each trial – in the 1-6s window following the onset of the textual cue and in the 1-6s window following the onset of the (neutral) facial expression. The following metrics from the Ledalab output were used in statistical analyses: (a) sum of CDA-identified SCR amplitudes within specified response window (i.e., CDA.AmpSum), (b) CDA-identified phasic driver within specified response window, which Ledalab developers recommend as it most accurately captures SCR (i.e., CDA.SCR), and (c) sum of TTP[39]-identified SCR amplitudes within specified response window (i.e., TTP.AmpSum). Specifically, these three metrics were standardised within each participant and included in three separate linear mixed models as dependent variables (Ben-Shakhar, 1985). Linear mixed models capturing SCR triggered by the textual cue included targets' group category membership (levels: ingroup member, outgroup member) and coalitional relations (levels: competitive, cooperative) as fixed effects and by-trial intercepts as a random effect[40]. Linear mixed models capturing SCR triggered by the facial expression followed the same structure with the only difference being the inclusion of emotion category (levels: angry, ambiguous, happy) as the third fixed effect of interest. Session ID (levels: first, second, third), which codes for the first, second and third batch of 32

[39] Trough-to-peak (TTP) scoring is the traditional and the most frequently adopted approach to quantify skin conductance response (Kuhn et al., 2022). In essence, TTP scoring involves calculating the difference between the skin conductance response at its maximum value (i.e., peak) and at its lowest preceding point (i.e., trough) within the time window prespecified by the researcher (1-6s in current research). TTP scoring has until recently mostly been done manually, but new softwares such as Ledalab (Benedek & Kaernbach, 2010) and Autonomate (Green et al., 2014) have fully automatised this quantification process.

[40] By-participant random intercepts were not included given that the three SCR metrics were standardised within each participant.

trials[41] (e.g., the first session captures trials 1-32), was included in all models to statistically control for expected habituation effects. This was done instead of restricting the SCR analyses to a specific section of the emotion task (e.g., the first session) because the four experimental conditions were balanced only on the level of the whole emotion task, not within each session.

Results

Gaze patterns. Analysis of average duration of participants' fixations revealed a significant main effect of AOI ($F(2, 35.4) = 27.95$, $p < .001$). Holm-corrected pairwise comparisons showed that observers fixated longer on the eye region of different facial expressions ($M = 405.89$, $SD = 237.75$), followed by fixations to nose region ($M = 373.72$, $SD = 249.36$, eyes-nose pairwise comparison: $z = 1.95$, $p = .052$) and the mouth region ($M = 228.87$, $SD = 294.55$, nose-mouth pairwise comparison $z = 5.67$, $p < .001$, eyes-mouth pairwise comparison: $z = 7.46$, $p < .001$). More importantly for the current set of research questions, I did not find a significant main effect of targets' group category membership (all *ps* across models with differently determined AOIs > .136), coalitional relations (all *ps* across models with differently determined AOIs > .209), or any interactions including these variables (all *ps* across models with differently determined AOIs > .150), regardless of how AOIs were defined.

Turning to dwell time, I again only observed a significant main effect of AOI (χ^2 (2) = 881.05, $p < .001$). Holm-corrected pairwise comparisons revealed that participants spent most time looking at the eye region regardless of the facial expression in question ($M = .61$, $SD = .21$), followed by the nose region ($M = .28$, $SD = .20$, eyes-nose pairwise comparison: *t*

[41] As the 2020 lab-based participants completed 144 trials of the emotion task instead of 96, in their case, variable session coded for the first, second and third batch of 48 trials.

= 56.63, $p < .001$) and the mouth region ($M = .11$, $SD = .15$, nose-mouth pairwise comparison: $t = 48.82$, $p < .001$, eyes-mouth pairwise comparison: $t = 97.35$, $p < .001$). Additionally, a trending main effect of condition (χ^2 (1) = 3.67, $p = .055$) and a trending two-way interaction between condition and AOI were observed (χ^2 (2) = 4.70, $p = .095$), with pairwise comparisons suggesting that participants looked marginally longer at the eyes of targets with whom they were to cooperate ($M = .62$, $SD = .21$) than at the eyes of targets with whom they were to compete ($M = .61$, $SD = .22$, $t = 1.73$, $p = .067$), while no such differences were observed in the nose region ($t = 0.39$, $p = .696$) or the mouth region ($t = -0.91$, $p = .363$). However, given that these effects were not even trending in models with alternatively defined AOIs (all *ps* for the main effect > .515, all *ps* for the interaction > .789), this pattern should not be overinterpreted. No other effects were observed across models (all *ps* > .101).

Analysis of saccade proportions revealed only a main effect of saccade type (χ^2 (2) = 161.90, $p < .001$), such that across all trials, the proportion of both intra-featural ($M = .38$, $SD = .20$) and inter-featural saccades ($M = .39$, $SD = .19$, intra-inter pairwise comparison: $t = 1.52$, $p = .129$) tended to be higher than the proportion of out-of-AOI saccades ($M = .23$, $SD = .22$, intra-out pairwise comparison: $t = 36.00$, $p < .001$; inter-out pairwise comparison: $t = 37.50$, $p < .001$), a pattern that is a direct consequence of how AOI radii were chosen, and is thus not particularly informative[42]. Importantly, regardless of how AOIs were defined, none of the main effects of emotion category (all *ps* > .472), targets' group category membership (all *ps* > .249), coalitional relations (all *ps* > .237), or the two-way and three-way interaction effects (all *ps* > .063) were significant. Finally, given that I found no robust evidence that group category membership, expected coalitional relations or emotion category had an

[42] For instance, when AOIs were narrowed down such that more fixations were not assigned to one of the three main AOIs (eyes, nose, mouth), the proportion of out-of-AOI saccades was unsurprisingly higher than the proportion of intra-featural and inter-featural saccades.

influence on visual attention allocation, the relationship between the analysed metrics and group, coalition, and emotion effects on judgments of facial expressions was not further explored.

Pupil responses. The results in the domain of pupil responses remained the same regardless of whether and how luminance was controlled for. Below, I report the output of the analyses which controlled for average luminance of textual cues and the local (i.e., mouth region) luminance of facial expressions. Examining pupillary responses following the onset of textual cues revealed no main effect of targets' group category membership ($p < .05$ for only 93.38 contiguous milliseconds), no main effect of coalitional relations with the target ($p < .05$ for only 40.02 contiguous milliseconds) and no interaction between the two factors ($p < .05$ for only 26.68 contiguous milliseconds) during the entire pupillary time-course. When examining pupillary responses following the presentation of targets' facial expressions, however, both the main effect of emotion category and the two-way interaction of targets' group category membership and expected coalitional relations met the duration criterion (200 ms) for statistical significance. Unpacking the main effect of emotion category (see Figure 2.10A), in the periods between 1995 and 2407 ms ($F_{AVERAGE} = 6.64$; $p_{AVERAGE} = .012$[43]), 2620 and 3910 ms ($F_{AVERAGE} = 6.99$; $p_{AVERAGE} = .007$), and 4043 and 5500 ms ($F_{AVERAGE} = 9.49$; $p_{AVERAGE} = .004$), participants reliably showed greater pupil dilation in response to angry expressions ($M_{1995\text{-}2407\ ms} = 1.18$, $SD_{1995\text{-}2407\ ms} = 2.19$; $M_{2620\text{-}3910\ ms} = 0.89$, $SD_{2620\text{-}3910\ ms} = 2.45$; $M_{4043\text{-}5500\ ms} = 0.77$, $SD_{4043\text{-}5500\ ms} = 2.74$), followed by ambiguous expressions ($M_{1995\text{-}2407\ ms} = 1.10$, $SD_{1995\text{-}2407\ ms} = 2.10$; $M_{2620\text{-}3910\ ms} = 0.80$, $SD_{2620\text{-}3910\ ms} = 2.45$; $M_{4043\text{-}5500\ ms} = 0.69$, $SD_{4043\text{-}5500\ ms} = 2.66$) and happy expressions ($M_{1995\text{-}2407\ ms} = 0.95$, $SD_{1995\text{-}2407\ ms} = 2.25$; $M_{2620\text{-}}$

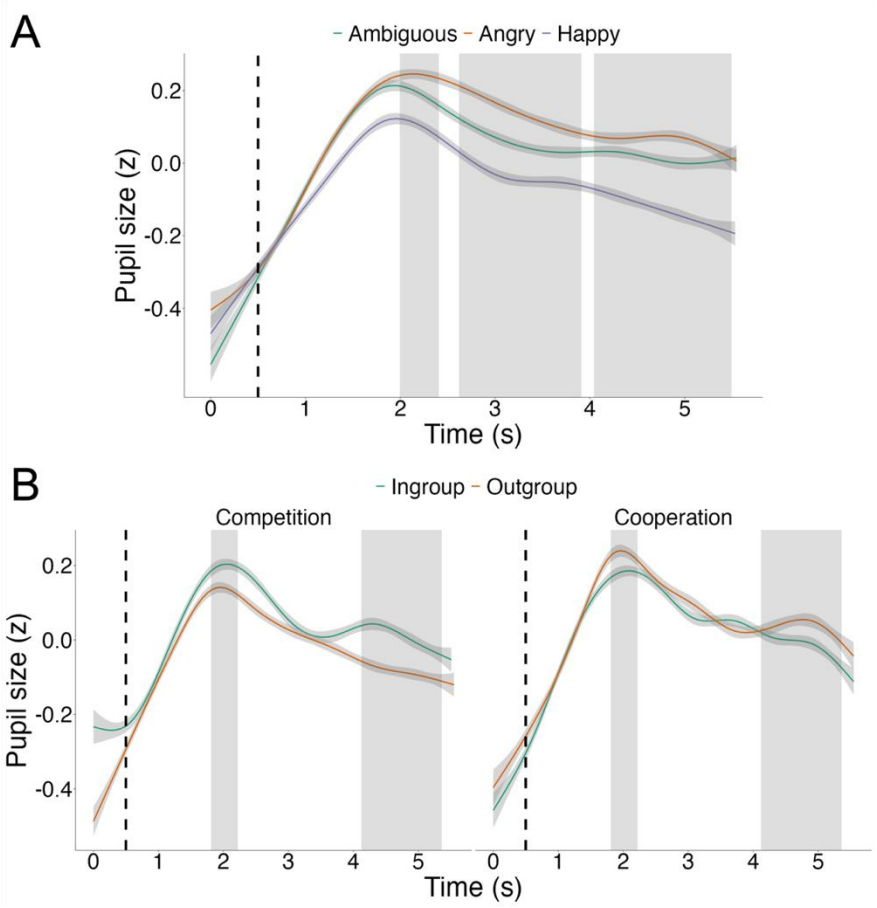
[43] For brevity, instead of reporting individual F statistics and p values from multiple (i.e., over 15) linear mixed models covering contiguous 13.34-ms windows of pupil responses, I report F statistics and p values averaged across these linear mixed models.

$_{3910\ ms}$ = 0.63, $SD_{2620\text{-}3910\ ms}$ = 2.47; $M_{4043\text{-}5500\ ms}$ = 0.44, $SD_{4043\text{-}5500\ ms}$ = 2.76[44]) with all Holm-corrected pairwise comparisons significant at $p < .001$. Unpacking the group category membership x coalitional relations interaction (see Figure 2.10B), in the periods between 1809 and 2221 ms ($F_{AVERAGE}$ = 7.91; $p_{AVERAGE}$ = .011) and between 4123 and 5399 ms ($F_{AVERAGE}$ = 7.82; $p_{AVERAGE}$ = .010) participants showed greater pupil dilation in response to expressions of ingroup members with whom they were to compete ($M_{1809\text{-}2221\ ms}$ = 1.15, $SD_{1809\text{-}2221\ ms}$ = 2.05; $M_{4123\text{-}5399\ ms}$ = 0.70, $SD_{4123\text{-}5399\ ms}$ = 2.68) and expressions of outgroup members with whom they were to cooperate ($M_{1809\text{-}2221\ ms}$ = 1.16, $SD_{1809\text{-}2221\ ms}$ = 2.09; $M_{4123\text{-}5399\ ms}$ = 0.72, $SD_{4123\text{-}5399\ ms}$ = 2.70) in comparison to expressions of ingroup members with whom they were to cooperate ($M_{1809\text{-}2221\ ms}$ = 1.06, $SD_{1809\text{-}2221\ ms}$ = 2.24; M = 0.62$_{4123\text{-}5399\ ms}$, $SD_{4123\text{-}5399\ ms}$ = 2.80) and expressions of outgroup members with whom they were to compete ($M_{1809\text{-}2221\ ms}$ = 1.02, $SD_{1809\text{-}2221\ ms}$ = 2.24; $M_{4123\text{-}5399\ ms}$ = 0.51, $SD_{4123\text{-}5399\ ms}$ = 2.70). In other words, in the periods between 1809 and 2221 ms and between 4123 and 5399 ms, participants showed greater pupil dilation to targets reflecting "incongruent" combinations of group category membership and coalitional relations (i.e., ingroup-competition, outgroup-cooperation) than to targets reflecting "congruent" combinations of group category membership and coalitional relations (i.e., ingroup-cooperation, outgroup-competition).

Figure 2.10

Changes in pupil size during the presentation of facial expressions

[44] These values capture means and standard deviations of pupil responses in relevant time intervals (e.g., between 4043 and 5500 ms).



Notes. Vertical dashed lines indicate when the neutral facial expression switched to an emotional facial expression. Shaded areas indicate the time intervals where the main effect of emotion category (A) and the two-way group category membership x coalitional relations interaction (B) were significant.

Skin conductance response (SCR). Analysis of all three metrics calculated as indices of SCR following the presentation of the textual cue yielded similar results. Specifically, regardless of which SCR metric was analysed, I observed either a significant or a trending main effect of coalitional relations (CDA.AmpSum model: $F(1, 3203.9) = 3.34$, $p = .067$; CDA.SCR model: $F(1, 3199.8) = 3.26$, $p = .070$; TTP.AmpSum model: $F(1, 2077.52) = 4.38$, $p = .036$), such that participants responded with greater physiological arousal to textual cues introducing targets with whom participants were to compete (CDA.AmpSum: $M = 0.03$, $SD = 1.05$; CDA.SCR: $M = 0.03$, $SD = 1.04$; TTP.AmpSum: $M = 0.04$, $SD = 1.09$) as opposed to cooperate (CDA.AmpSum: $M = -0.03$, $SD = 0.94$; CDA.SCR: $M = -0.03$, $SD = 0.94$; TTP.AmpSum: $M = -0.04$, $SD = 0.89$) in the second part of the experiment. None of the

models revealed a significant main effect of targets' group category membership (all *ps* > .344) or a significant two-way interaction (all *ps* > .795).

Turning to participants' SCR following the presentation of the targets' facial expressions, the analysis of the three metrics yielded somewhat dissimilar results and for this reason, I will present the results from each of the three models separately. First, the model with CDA-identified SCR amplitudes (i.e., CDA.AmpSum in Ledalab) as the dependent variable revealed a significant main effect of coalitional relations ($F(1, 2808.94) = 5.29$, $p = .021$), such that participants' SCR was significantly stronger when participants were exploring the facial expressions of targets with whom they were to compete ($M = 0.03$, $SD = 1.04$) as opposed to cooperate ($M = -0.03$, $SD = 0.95$) in the future. In this model, I also observed a trending group category membership x coalitional relations x emotion category interaction ($F(2, 2805.87) = 2.55$, $p = .078$), which suggested that the pattern of SCR upon the presentation of ambiguous and angry facial expressions could be better characterised by independent contributions of targets' group category membership and coalitional relations while the pattern of SCR following the presentation of happy facial expressions could be better characterised by an interaction between these two factors. However, this pattern of mean values was observed in the absence of any significant pairwise comparisons (all *ps* > .088) and should thus not be overinterpreted. No other significant effects were observed (all *ps* > .123).

Second, the model with CDA-identified phasic activity (i.e., CDA.SCR in Ledalab) as the dependent variable revealed both a trending main effect of targets' group category membership ($F(1, 2809) = 3.66$, $p = .056$), such that participants responded with greater physiological arousal to ingroup targets' expressions ($M = 0.03$, $SD = 1.09$) than outgroup targets' expressions ($M = 0.03$, $SD = 0.88$), and a trending main effect of coalitional relations with the target ($F(1, 2809) = 3.62$, $p = .057$), such that stronger SCR was observed while

participants were exploring the facial expressions of targets with whom they expected to compete (M = 0.03, SD = 1.01) than to cooperate (M = -0.03, SD = 0.98). No group category membership x coalitional relations x emotion category interaction was observed (F(2, 2809) = 1.79, p = .168), and nor were any other significant effects in the model (all ps > .242).

The third and final model with TTP-identified SCR amplitudes (i.e., TTP.AmpSum in Ledalab) as the dependent variable revealed a significant main effect of targets' group category membership (F(1, 1313) = 4.91, p = .027), indicating that increased SCR occurred in response to ingroup targets' facial expressions (M = 0.04, SD = 1.13) compared to outgroup targets' facial expressions (M = -0.05, SD = 0.83). I again found a trending group membership x coalitional relations x emotion category interaction (F(2, 1313) = 2.45, p = .087), suggesting the same pattern of relations between the three variables as the CDA.AmpSum model, but this interaction again did not yield any significant pairwise comparisons (all ps > .248) and will not be further interpreted. Given that I observed the same pattern of effects on SCR and valence judgments of facial expressions, I proceeded to analyse the correlations between participants' SCR metrics and their corresponding valence judgments. These analyses are reported in Appendix 2.1. For transparency, it is important to note that at no point was a direct association between physiological responses and participants' valence judgments of facial expressions anticipated, and for this reason, the analyses in Appendix 2.1 should be treated as exploratory.

Discussion

Experiments 1 and 2 provided evidence that participants judged valence of facial expressions differently depending on the group category membership of the expresser and the coalitional relations with the expresser. Experiment 3 sought to explore whether more covert processing of facial expressions is similarly affected by group category membership and coalitional relations. Specifically, I investigated whether the influence of group category

membership and coalitional relations is restricted to how facial expressions are interpreted, or whether these factors penetrate the level of (a) visual attention, influencing how people attend to facial features, and (b) physiological arousal, reflecting perceivers' physiological responsivity to incongruent combinations of social information about the expresser (e.g., incongruent pairings of group category membership and coalitional relations).

In the domain of visual attention, Experiment 3 found no evidence that knowing the expresser's group category membership and whether one would interact cooperatively or competitively with them changed how participants subsequently attended to expresser's facial features. These results are in line with the original interpretation offered by Lazerus and colleagues (2016) and suggest that ingroup positivity bias and cooperation positivity bias in judgments of facial expressions are more likely a consequence of a spillover of positivity from ingroup and cooperation to the domain of corresponding judgments of facial expressions than a consequence of any group- or coalition-tailored visual attention mechanisms.

In the domain of physiological arousal, Experiment 3 showed that expresser's group category membership and expected coalitional relations reflected in participants' skin conductance response (SCR) and pupil size, but the results were somewhat inconsistent across the two measures and different SCR metrics. First, during the presentation of textual cues informing participants of the upcoming expresser's group category membership and coalitional relations with them, I observed the following: (a) stronger SCR when the textual cue introduced an expresser with whom participants were to compete as opposed to cooperate in the future, and (b) no differences in pupil size regardless of which expresser the textual cue introduced (e.g., ingroup or outgroup member with whom the participant was to compete or cooperate). These results do not support the initial hypothesis that participants' physiological arousal would track group-coalition incongruences in the textual cues but can still be

interpreted in the context of previous research showing that competitive interactions cause an emotional state characterised by high arousal (Scott & Cherrington, 1974; Ku et al., 2005; To et al., 2018). Given that competition can be appraised as both a challenge and a threat (Moore et al., 2013; To et al., 2020) and that SCR primarily signals arousal, and not valence of the emotional state (Anders et al., 2004; Bell et al., 2018), it is not possible to say whether participants had a positive (e.g., excitement) or a negative (e.g., stress) reaction to "competitive" textual cues. Finally, due to this established link between arousal and competition in the literature, it is somewhat surprising that "competitive" textual cues did not similarly elicit increased pupil dilation. However, while pupil dilation tends to correlate with SCR (Bradley et al., 2008; Schumann et al., 2020; repeated-measures correlation in Experiment 3: r = .06-.08[45], all ps < .001), it is not unusual for these two signals to show different patterns of results (see Kinner et al., 2017; Leuchs et al., 2019 for examples). Nonetheless, one experiment-specific detail should be noted; in Experiment 3, participants were not freely exploring the textual cues, but were tasked to memorise the information contained in them to subsequently answer two attention checks. Memorising repetitive information could have on its own led to increased pupil dilation via increased cognitive effort and thus potentially restricted my ability to detect effects stemming from the exact content of textual cues. This limitation might similarly apply to SCR which has also been used to index information processing and cognitive effort in the past (Dawson et al., 1989; Filion et al., 1991; Naccache et al., 2005). Nonetheless, given that there is no reason to assume that "competitive" textual cues were more difficult to memorise than "cooperative" textual cues, differential cognitive effort is unlikely to account for observed SCR effects, rendering the association between arousal and (anticipated) competition the best candidate interpretation. Overall, it remains an open question whether textual cues provided an optimal

[45] Multiple correlation coefficients are calculated due to the existence of multiple SCR metrics.

way to detect group- and coalition-triggered changes in physiological arousal, and alternative experimental set-ups which do not involve co-occurring working memory load should be explored in future research.

Second, during the presentation of facial expressions of emotion, I observed the following: (a) stronger SCR when participants were viewing expressions of minimal ingroup relative to minimal outgroup members, and (b) stronger SCR when participants were viewing expressions of others linked to a competitive as opposed to a cooperative context, with both effects varying across the three included SCR metrics. Furthermore, I observed greater pupil dilation (a) when participants were viewing angry facial expressions in comparison to ambiguous facial expressions and ambiguous facial expressions in comparison to happy facial expressions, and (b) when participants were viewing facial expressions of "incongruent" expressers (i.e., ingroup members with whom participants were to compete and outgroup members with whom participants were to cooperate) relative to "congruent" expressers (i.e., ingroup members with whom participants were to cooperate and outgroup members with whom participants were to compete). Given that emotion category is the only factor that was acutely salient during the presentation of facial expressions, I initially expected SCR and pupil dilation to reflect emotion-group and/or emotion-coalition incongruences (Gamond et al., 2017). This hypothesis was not supported, but the patterns of pupil dilation can instead be considered to reflect group-coalition incongruence. Specifically, the candidate interpretation of the latter pupil dilation effect, as suggested in its description, might be that participants showed higher levels of physiological arousal because viewing incongruent group-coalition pairings runs counter to their expectation about how minimal ingroup and outgroup members should interact within and across group boundaries. While SCR did not similarly track group-coalition incongruence, it should be noted that independent influences of group category membership and coalitional relations imply greatest SCR in

response to minimal ingroup members with whom one is to compete and lowest SCR in response to minimal outgroup members with whom one is to cooperate, which could signal that participants were more responsive to violations of within-group than between-group expectations. Finally, while I was not theoretically interested in whether physiological arousal varies with different emotion categories, the finding that pupils dilated most in response to angry facial expressions adds to the previous literature which suggested that these expressions might be particularly arousing because they serve as a signal of interpersonal threat (Kret et al., 2013).

General discussion

In this chapter, I sought to disentangle the influence of minimal group membership and coalitional relations on valence, arousal, and genuineness judgments of facial expressions of emotion. Experiments 1 and 2 documented that valence judgments are indeed biased by expresser's minimal group membership and anticipated coalitional relations with them, while arousal and genuineness judgments never showed any signs of group- or coalition-related bias. Specifically, Experiments 1 and 2 provided support for two independent biases: (a) an *ingroup positivity bias*, wherein minimal ingroup members' facial expressions were judged as conveying more positive emotions than minimal outgroup members' facial expressions, and (b) a *cooperation positivity bias*, wherein facial expressions of those linked to a cooperative context were judged as conveying more positive emotions than facial expressions of those linked to a competitive context.

These findings expand on previous minimal group studies which similarly observed ingroup positivity bias in judgments of facial expressions. However, these previous studies either turned minimal groups against each other (e.g., Lazerus et al., 2016), thus confounding minimal group membership with coalitional relations, or failed to discuss the fact that within-group cooperation is automatically assumed (e.g., Dunham, 2011). This state of the literature

made it difficult to discern whether ingroup positivity bias was indeed driven by minimal group membership as such, rendering the name of this bias justified, or by explicitly or implicitly co-occurring coalitional relations, rendering cooperation positivity bias a more appropriate name. By treating minimal group membership and coalitional relations as orthogonal, experiments described in the current chapter disambiguated past work and showed that both factors independently influence valence judgments of facial expressions. Thus, if these findings were to be translated to a real-life scenario, one might hypothesise that a prototypical intergroup conflict environment, where ingroup members are perceived as cooperative and outgroup members as competitive, might be the context in which people are most likely to succumb to bias in valence judgments of facial expressions.

Nevertheless, before engaging in any translational discussions, it is important to reflect on the practical significance of results from Experiments 1 and 2 in addition to their statistical significance, especially given that these experiments are highly powered and p values change as a function of sample size (Sullivan & Feinn, 2012). Even though linear mixed models, used in Chapter 2, do not allow calculation of standardised effect sizes (e.g., partial eta squared) for individual fixed effects (Nakagawa & Cuthill, 2007), examining unstandardised mean differences reveals that ingroup positivity bias and cooperation positivity bias are in fact very small. Specifically, across analyses in Experiments 1 and 2, ingroup positivity bias was reflected in mean differences between 0.68 and 3.71 on a 100-point scale while cooperation positivity bias was reflected in mean differences between 0.43 and 1.15 on the same scale. These mean differences are comparable to those reported in the previous literature (Dunham, 2011; Lazerus et al., 2016) and are not surprising given the

weak experimental manipulations[46], but they still caution against treating minimal group membership and coalitional relations as sources of major influence on judgments of facial expressions of emotion. I would argue that the meaningfulness of ingroup and cooperation positivity biases observed in Experiments 1 and 2 lies primarily in their ability to facilitate understanding of previous research documenting similarly sized minimal group effects on judgments of facial expressions, but it should be noted that other factors (e.g., body posture, situational context, Hassin et al., 2013) are likely to influence judgments of facial expressions more substantially. Finally, the same reasoning extends to the effects observed in the domain of peripheral physiology (discussed in more detail below). While skin conductance response and pupil dilation have the potential to reveal implicit processing steps that do not necessarily manifest in explicit judgments, the effects of minimal group membership and coalitional relations are substantially smaller than the effects of more arousing content (Bradley & Lang, 2007) and low-level physical properties (e.g., luminance; Cuve, Stojanov et al., 2021).

Furthermore, whereas Experiments 1 and 2 only focused on final judgments of facial expressions, which offer little insight into real-time emotion processing, Experiment 3 adopted eye-tracking to gain a more mechanistic understanding of how social information about the expresser may drive the processing of their facial expressions. In particular, I examined whether ingroup and cooperation positivity biases could be traced back to differential patterns of visual attention paid to expressers' facial features. Experiment 3 found no evidence in support of this possibility. Specifically, expresser's group category membership and expected coalitional relations with them did not affect average duration of individual fixations on different facial features (i.e., eyes, nose, mouth), total viewing time of

[46] Experiments 1 and 2 randomly assigned participants to minimal groups instead of using bogus personality tests or art preferences. Similarly, these experiments relied only on the *anticipation* of coalitional relations, without having participants actually interact competitively and cooperatively with others.

these facial features or the number of saccades within and between facial features. This means that minimal group membership and coalitional relations are unlikely to have exerted influence on valence judgments through distinct, top-down attentional shifts, rendering the original hypothesis offered by Lazerus and colleagues (2016) a better candidate explanation for observed effects. Specifically, Lazerus and colleagues (2016) proposed that ingroup positivity bias likely arises post-perceptually because the positivity associated with ingroups exerts top-down influence on corresponding judgments of facial expressions, akin to the classical idea of affective priming (Fazio, 2001; Houwer, 2009). The current chapter expands on this hypothesis to suggest that positivity might not be restricted to ingroups but might also be associated with cooperation, thus allowing both factors to exert top-down positivity effects.

Of note, the current experimental design does not allow me to distinguish whether minimal group membership and coalitional relations influenced valence judgments by making ingroup members' and cooperative others' facial expressions seem more positive or by making outgroup members' and competitive others' facial expressions seem more negative. In other words, based on current experiments alone, it remains unclear whether the observed effects stem from the positivity associated with ingroups and cooperation or from the negativity associated with outgroups and competition. Past research suggested that ingroup-favouring biases in minimal intergroup contexts reflect own-group love (i.e., positive associations with and feelings towards one's own group), rather than other-group hate (i.e., negative associations with and feelings towards the other group, Molenberghs et al., 2014; Molenberghs & Louis, 2018; see also McAuliffe & Dunham, 2016), making it more likely that ingroup positivity bias indeed stems from the positivity associated with ingroups rather than from the negativity associated with outgroups. However, this distinction is less clear in the domain of coalitional relations. Thus, future research should assess whether *cooperation*

positivity bias is true to its name or could be better understood as *competition negativity bias*. To achieve this, future research could adopt an alternative experimental set-up where participants first judge facial expressions of unaffiliated expressers, in the absence of information about minimal group membership and coalitional relations, and then proceed to judge the same facial expressions again, but this time in the presence of group- and coalition-related information. This design would help disentangle whether the observed coalition effect reflects cooperation positivity bias or competition negativity bias by examining how valence judgments change in the presence of anticipated competition and cooperation in comparison to baseline valence judgments of unaffiliated others' facial expressions.

Another characteristic of the current experiments is that competition and cooperation were features of the social context rather than a result of anyone's choice. This means that participants thought that ingroup and outgroup members did not choose to compete or cooperate but were instead randomly assigned to a competitive or a cooperative context. For this reason, minimal ingroup members could not have been perceived as intentionally violating the within-group cooperation norm. Future research should thus investigate to what extent the current findings generalise to ingroup members who deliberately deviate from the expectation to cooperate instead of simply finding themselves in a competitive intragroup context. My prediction is that norm-violating ingroup members would be evaluated more negatively than ingroup members who are randomly assigned to a competitive context (Mendoza et al., 2014; Wang et al., 2016), resulting in former ingroup members' facial expressions being judged as conveying more negative emotions than latter ingroup members' facial expressions.

In addition to eye-tracking, Experiment 3 adopted two measures of peripheral physiology – pupil dilation and skin conductance response (SCR) – to index implicit processing of social information about expressers. With the help of these measures,

Experiment 3 documented that participants showed higher physiological arousal, as indexed by increased SCR, in response to anticipated competition and ingroup members versus anticipated cooperation and outgroup members. Simultaneously, people showed higher physiological arousal, as indexed by increased pupil dilation, in response to incongruent versus congruent group-coalition pairings (see also Hepach & Westermann, 2013; Proulx et al., 2017). To unpack this result, participants displayed increased pupil dilation in response to facial expressions of ingroup members with whom they were to compete and outgroup members with whom they were to cooperate (i.e., incongruent group-coalition pairings) than to facial expressions of ingroup members with whom they were to cooperate and outgroup members with whom they were to compete (i.e., congruent group-coalition pairings). Even though patterns of SCR did not completely align with patterns of pupil dilation, independent influences of minimal group membership and coalitional relations on SCR still imply that participants showed the highest physiological arousal in response to facial expressions of ingroup members with whom one was to compete. Therefore, both measures agree that participants displayed higher physiological arousal in response to violations of within-group coalitional expectations, which aligns with previous literature suggesting that minimal ingroup members are more deterministically expected to cooperate than minimal outgroup members are expected to compete (Balliet et al., 2014; Brewer, 2008; Yamagishi & Kiyonari, 2000, see also Puurtinen et al., 2015). From here, future research could examine whether participants who more strongly associate "ingroups" with "cooperation" and "outgroups" with "competition" show greater physiological arousal to violations of group-coalition pairings in comparison to participants who more weekly associate these concepts. If this is confirmed, it would suggest that measures of peripheral physiology, in addition to mouse-tracking (see Stolier & Freeman, 2016) and neural measures (see Gamond et al., 2017), can

provide insight into individual perceivers' lay beliefs about how different social phenomena correlate.

Finally, even though we can experimentally distinguish group category membership from coalitional relations – making people positively interdependent (i.e., cooperative), independent, or even negatively interdependent (i.e., competitive) in a way that aligns or misaligns with minimal group membership – is doing this representative of how groups operate? In other words, while one can arguably recall a scenario where members of two different groups were competitive or cooperative with each other, rendering intergroup relations variable in terms of their coalitional implications, can one imagine ingroup members being anything other than cooperative with each other? This question boils down to: what is the reason to experimentally disentangle groups from cooperation?

Given that cooperation normatively co-occurs with sharing a group category membership, it can be difficult to discern how big of a role it plays in groups' influence on people's judgments, feelings, and behaviours. Explicitly outlining the role of cooperation behind minimal group effects is particularly important in light of coalitional accounts of the minimal group paradigm, which suggest that minimal groups serve as a signal of who is likely to be a preferential cooperator (Yamagishi & Kiyonari, 2000). According to these accounts, minimal ingroup biases emerge due to individuals' expectations of within-group interdependence and cooperation (Cikara, 2020; Misch et al., 2021; Pietraszewski, 2013; Yamagishi & Kiyonari, 2000; see also Powell, 2022). Therefore, if one were to test whether (anticipated) cooperation underlies a certain minimal ingroup bias – for example, ingroup positivity bias in judgments of facial expressions – a useful approach would be to decorrelate (anticipated) cooperation from group category membership. If a minimal ingroup bias of interest disappears in the absence of (anticipated) cooperation, then (anticipated) cooperation can be considered an important driver of what was originally thought to be a black-box group

category membership effect (see Misch et al., 2021). Connecting this discussion to Experiments 1-3, if (anticipated) cooperation completely accounted for ingroup positivity bias in judgments of facial expressions, participants would switch to rating cooperative others' facial expressions as conveying more positive emotions than competitive others' facial expressions, while no effects of minimal group membership would remain. This is, however, not what Experiments 1-3 found as minimal group membership continued to exert its influence on judgments of facial expression even when minimal ingroup members did not serve as a reliable signal of potential cooperation. In summary, the primary reason to disentangle minimal groups from (anticipated) cooperation, as well as other relevant features such as intrinsic similarity, is to test the extent to which (anticipated) cooperation and other features underlie the formation of a variety of ingroup biases in minimal groups.

To conclude, the observed influences of minimal group membership and coalitional relations on valence judgments of facial expressions indicate that previously documented ingroup positivity bias cannot be fully accounted for by explicitly or implicitly co-occurring coalitional relations. Based on current results, it seems that both ingroup and cooperation positivity biases are best understood as examples of affective priming, where the evaluative response to the idea of shared group membership and cooperative relations with the expresser translates to the domain of valence judgments of facial expressions. Finally, Experiments 1-3 illustrated how peripheral physiology can index steps in implicit processing of social information, which were not evident in participants' explicit responses. Given that measures of peripheral physiology tend to be more widely accessible and easier to set up in comparison to neural measures, future research should continue to explore the potential of these measures to offer insights into various aspects of social information processing.

CHAPTER 3: PERCEPTIONS OF EMPATHIC NORMS IN INTERGROUP CONTEXTS

Abstract

Groups are known to influence empathy – people empathise less with outgroup members than with ingroup members, or in more extreme cases, even take pleasure in outgroup suffering. One promising way to encourage empathy for outgroups involves portraying intergroup empathy as normative, but little is currently known about people's unaltered perceptions of these *empathic norms* in intergroup contexts, which arguably provide the basis for most normative effects outside the laboratory. Therefore, in three experiments, spanning ethnic, political, and minimal intergroup contexts, I sought to provide a first comprehensive description of these subjective perceptions, their accuracy, and dependency on contextual information about empathic targets and their emotional experiences as well as on individual differences in how sampled group relations are perceived. The results revealed that perceptions of empathic norms are overly negative, with people expecting more intergroup empathy bias from others than they themselves showed and overestimating the bias actual ingroup and outgroup members showed. In competitive real-world, but not minimal intergroup contexts, I found evidence of *empathic norm perception asymmetry*, wherein outgroup members were assumed to discriminate more than ingroup members in their empathy, especially when empathic targets' social identities were salient. Finally, perceptions of empathic norms were reliably predicted by corresponding perceptions of prejudice, pointing to the possibility that people rely on their general representations of group relations to make specific social predictions in novel circumstances. Taken together, these findings document people's tendency to overestimate intergroup antagonism even in the domain of empathy, which is seen as moral and socially desirable,

and highlight the need to expand the focus of current models of intergroup social perception to include representations of group relations in addition to groups' traits.

Introduction

Chapter 1 reviewed research concerning the influence of groups on *empathy*, which, in current work, is defined as vicarious adoption of the empathic target's emotions in terms of valence. This literature has revealed a strikingly robust *intergroup empathy bias*, wherein people empathise more with ingroup members than outgroup members (Bruneau et al., 2015; Bruneau et al., 2017; Montalan et al., 2012; Vollberg et al., 2021), and has documented a vast array of its consequences, ranging from intergroup neglect to intergroup aggression (Cikara, 2015). Research has further suggested that empathy for outgroup members can be increased when intergroup empathy is portrayed as normative. For instance, people empathise more with outgroup members when ingroup members are also empathetic towards the same outgroup (conformity to *ingroup intergroup empathic norms*, Saguy et al., 2015; Tarrant et al., 2009, see also Nook et al., 2016) or when outgroup members are empathetic towards the ingroup (reciprocity of *outgroup intergroup empathic norms*, Gubler et al., 2015; Hein et al., 2016; Nadler & Liviatan, 2006). There is some variability in how previous research defined and operationalised intergroup empathic norms, but in current work, *intergroup empathic norms* refer to social norms which signal how much feelings of outgroup empathic targets are vicariously adopted in one's reference group, akin to descriptive rather than injunctive social norms in Cialdini's (2003) terminology[47].

Previous findings are illustrative of intergroup empathy's susceptibility to normative influence, but it is important to note that previous research has been homogeneous in its reliance on experimental manipulations of intergroup empathic norms. Needless to say,

[47] Previous research has found that people tend to judge what is common as also appropriate and desirable (Lindström et al., 2018; Tworek & Cimpian, 2016), making it likely that perceptions of descriptive and injunctive intergroup empathic norms will be aligned. However, given that Experiments 4, 5, and 6a explicitly assessed only descriptive intergroup empathic norms, in the absence of additional data on injunctive intergroup empathic norms, I restrict any interpretations to the descriptive domain.

people do not enter these experimental manipulations as blank slates, but presented normative information is assessed in light of people's pre-existing perceptions of intergroup empathic norms (Tankard & Paluck, 2016; van der Linden, 2013). If the distance between normative information contained in the experimental manipulations and people's subjective perceptions of the same norm is too large, that might reduce the manipulations' plausibility and thus, their effectiveness (Tankard & Paluck, 2016, see also Experiment 6b in Chapter 4). Therefore, before manipulating intergroup empathic norms in a specific intergroup context, it is crucial to get a sense of people's *unaltered* perceptions of these norms to achieve the optimal plausibility of subsequent experimental manipulations.

Thus far, I have focused on the interaction between deliberate manipulations and pre-existing perceptions of intergroup empathic norms, but more broadly, it is important to recognise that subjective perceptions of social norms are the primary channel of normative influence outside the research environment (Jachimowicz et al., 2018; Tankard & Paluck, 2016; Vallentin-Holbech et al., 2017). In other words, regardless of their potential inaccuracy, subjective perceptions of social norms represent a reality within which individuals operate. While experimental manipulations expose people to certain normative information, after which people's adaptation to this information is tested, in the real world, people often do not have any immediate cues of social norms but instead rely on their formed perceptions of these norms to guide their actions. In summary, despite the fact that subjective perceptions of intergroup empathic norms provide insight into individually perceived intergroup reality and despite the importance of these perceptions for designing believable empathy-building interventions, they are currently under-researched.

The only study that has refrained from manipulating empathic norms to examine their unaltered perception has documented that people tend to underestimate inclusive empathy – "empathy for one's fellow citizens affected by (…) conflict experiences regardless of their

(…) group belonging" (Penić et al., 2021, p. 402) – in their immediate communities (i.e., among neighbours). This study provides an excellent starting point for developing hypotheses about how perceived empathic norms (i.e., how much empathy relevant others are assumed to feel) compare to actual empathic norms (i.e., how much empathy relevant others actually feel). However, due to its exclusive focus on perceptions of ingroup empathic norms in a single intergroup environment (post-war Sri Lanka), it also leaves many questions unanswered.

This chapter thus picks up on Penić et al.'s (2021) research and reports three experiments spanning different intergroup contexts (Experiments 4, 5 and 6a) that address the following research questions: (a) how much empathy are average ingroup and outgroup members assumed to feel towards own-group versus other-group empathic targets (addressed in Experiments 4-6a); (b) how the levels of intergroup empathy bias participants assume that ingroup and outgroup members show compare to the levels of intergroup empathy bias participants report themselves (addressed in Experiments 4-5) and (c) to the levels of intergroup empathy bias ingroup and outgroup members in fact report on average (addressed in Experiments 4-5); (d) whether participants assume that average ingroup and outgroup members discriminate equally in their empathy (addressed in Experiments 4-6a); (e) whether the availability of different contextual information about empathic targets and their emotional experiences influences how much intergroup empathy bias average ingroup and outgroup members are assumed to show (addressed in Experiments 4-5); and (f) whether participants' general perceptions of intra- and intergroup relations predict their perceptions of empathic norms (addressed in Experiments 4-5).

This chapter is divided into three sections corresponding to Experiments 4-6a which were conducted in the following intergroup contexts: the post-conflict context of Croatia and Serbia (Experiment 4), the political context surrounding the UK's departure from the

European Union ("Brexit"; Experiment 5) and a minimal intergroup context (Experiment 6a). Having described each experiment and its results separately, I conclude with a general discussion, in which I consider how the findings from the three experiments advance existing research on empathic norms, suggest future directions, and discuss strengths and limitations of the present studies.

Chapter 3 closely follows Chapter 2 in its general effort to examine (perceptions of) group relations alongside group category membership and in its pursuit of identifying generalisable *features* of group environments that are likely to influence socio-affective processes, regardless of the exact groups under consideration. Nonetheless, as is likely evident from the introduction, these two chapters take different research approaches (i.e., manipulating intra- and intergroup relations in Chapter 2, measuring perceptions of intra- and intergroup relations in Chapter 3), focus on different socio-affective processes (i.e., emotion judgments in Chapter 2, empathy and its relation to empathic norms in Chapter 3), emotion modalities (i.e., facial expressions of emotion in Chapter 2, text-based expressions of emotion in Chapter 3) and group contexts (i.e., minimal groups in Chapter 2, a combination of minimal and real-world groups in Chapter 3)[48] and for this reason, Chapter 3 does not directly pick up on the findings presented in Chapter 2.

Hypotheses

In this section, I will briefly outline and justify hypotheses pertaining to each research question. Due to the number of inter-related research questions, these hypotheses will be repeated throughout the chapter, where deemed relevant.

[48] It is important to note that the COVID-19 pandemic also played a crucial role in this change of research focus as it made original research plans, which were heavily reliant on lab-based experiments and methods, unforeseeable at best and unfeasible at worst.

H1 (research questions a-c). People tend to view themselves as less susceptible to intergroup biases than others (Pronin, 2007), attribute more desirable emotions to themselves than others (Ong et al., 2018) and overestimate the antagonism of intergroup relations (Lees & Cikara, 2020; Moore-Berg et al., 2020; Ruggeri et al., 2021). Therefore, my first hypothesis was that participants would not only assume that others show intergroup empathy bias (research question a) but also that others show more intergroup empathy bias than participants do themselves (research question b). Further, I predicted that participants would tend to overestimate the extent to which actual ingroup and outgroup members show intergroup empathy bias (research question c, Penić et al., 2021).

H2 (research question d). Since equitable empathy is seen as morally superior to biased empathy for close others (Fowler et al., 2021) and ingroups tend to be assigned more positive characteristics, including morality, than outgroups (Enock et al., 2021; Liberman et al., 2018), it seemed possible that biases in perception of empathic norms are asymmetrical. Specifically, I assessed whether people assume that outgroup members discriminate more than ingroup members in their allocation of empathy.

H3 (research question e). Perceptions of empathic norms likely depend on available contextual information. Knowing specific empathic targets' personal experiences (e.g., what exactly happened to them and how they responded) should reduce the likelihood that these empathic targets are represented solely in terms of their social identities (Bruneau et al., 2015; Sheng & Han, 2012), facilitating perspective taking and empathy (Gaesser et al., 2020) and potentially reducing bias in both empathy and perceptions of empathic norms. The reason is that contextual information allows participants to take into account the idiosyncrasies of the empathic targets' emotional circumstances when making their judgments (Ong et al., 2018). Thus, when compared to *acontextualised* perceptions of empathic norms, reported in the absence of any individual empathic targets and emotional circumstances, *contextualised*

perceptions of empathic norms, reported in the context of individual empathic targets and emotional circumstances, should reduce assumptions of intergroup empathy bias (addressed in Experiment 4). However, these debiasing effects should only be observed when the contextual information draws attention towards each empathic target's unique emotional circumstances and away from their social identity (see Bruneau et al., 2015). By contrast, when contextual information instead emphasises the empathic target's social identity (for example, by revealing how a political outgroup member's personal emotional experience has shaped their political affiliation), any beneficial effect of contextual information might be lost (addressed by Experiment 5). Under these circumstances, the encoding of social identities is no longer disrupted, and representation of the empathic target in terms of their group category membership may even be enhanced by their emotional circumstances.

H4 (research question f). In line with research on interpersonal social cognition which has mapped how people use their general perceptions of other *individuals'* traits to infer individuals' mental states on specific occasions (Conway et al., 2020; Long et al., 2022; Tamir & Thornton, 2018), people might similarly use their general perceptions of *groups'* traits to infer group members' emotions and behaviour on specific occasions[49]. However, when trying to infer group members' emotions and behaviour in intergroup contexts as opposed to other occasions, as is the case in Experiments 4 and 5, it might be more diagnostic to rely on perceptions of relevant intra- and intergroup relations (e.g., whether the relations between groups in question are warm) instead of representations of groups' traits (e.g.,

[49] Please note that this marked distinction between "individuals" and "group members" is used only to bridge two lines of social psychological research more easily. In real life, the same social target can be viewed as an "individual" and as a "group member" allowing perceivers to use both their perceptions of the target's unique traits (e.g., "my colleague is a warm person"), and their perceptions of the target's social background and the traits associated with it (e.g., "my colleague comes from a country where people are warm") to make inferences about the target's mental states on specific occasions.

whether group members are warm in general). Given that a person's own experience of empathy is enhanced when empathic targets are liked (Gutsell & Inzlicht, 2010; Gutsell & Inzlicht, 2012; Marsh, 2022), the present research tested whether people's perceptions of group relations' valence (Tamir & Thornton, 2018), embedded in people's beliefs about how positively or negatively group members feel towards their own groups and each other (*perceptions of ingroup and outgroup prejudices*), inform their expectations regarding ingroup and outgroup members' empathy in novel circumstances. If this is the case, individual differences in prejudice perceptions should predict individual differences in perceptions of empathic norms. For instance, if people believe that outgroup members dislike or even hate the ingroup, this prejudice perception should specifically predict how much empathy they expect outgroup members to show for ingroup, but not outgroup targets. Please note that even though this section of Experiments 4 and 5 is correlational, my underlying assumption is causal – if I indeed observe a consistent pattern of associations between prejudice perceptions and perceptions of empathic norms, my assumption is that participants relied on their general perceptions of known group relations to generate their empathic predictions (see Grosz et al., 2020).

Hypotheses pertaining to emotion perception. Finally, although the research described in this chapter was designed to advance understanding of intergroup empathy with respect to perceptions of empathic norms, I also considered it important to include emotion perception measures when assessing empathy and perceptions of empathic norms in the context of specific emotional experiences (Coll et al., 2017). The reason is that empathy and perceptions of empathic norms are modulated by assessment of the empathic target's emotional state (Coll et al., 2017) – the more negatively perceivers judge the empathic target to be feeling, the more negatively they tend to feel themselves in response, and, by extension, the more negatively perceivers think that relevant others judge the empathic target to be

feeling, the more negatively they might expect those others to feel in response. Thus, to avoid conflating perception and sharing of emotions (Coll et al., 2017), I included emotion perception measures and examined whether participants also assume others would show intergroup bias in their perception of empathic targets' emotional states, perceiving own-group members' emotional states as more negative than equivalent other-group members' emotional states (Richins et al., 2019). However, as emotion perception does not have such strong connotations of morality as empathy does (Fowler et al., 2021), I did not expect to observe ingroup-favouring asymmetry in participants' expectations about emotion perception. In summary, measuring how ingroup and outgroup empathic targets' emotions are perceived and assumed to be perceived by others in addition to measuring how much their emotions are shared and assumed to be shared by others allowed me to delineate two related processes – emotion perception and empathy.

Experiment 4

Method

Intergroup context. Croatia and Serbia, modern-day countries in south-eastern Europe, were previously constituent states in the Socialist Federative Republic of Yugoslavia (SFRY). In 1991, Croatia declared its independence from SFRY but the ethnic Serb minority, which constituted 12.2% of the total population in Croatia at the time (Croatian Bureau of Statistics, 2005), rejected this declaration (UN-ICTY). The ethnic Serb minority continued to proclaim a Serbian autonomous region within Croatia, which led to a full-scale war between Croatian forces on one side and ethnic Serb forces, supported by the Yugoslav National Army and Serbia, on the other (UN-ICTY). The fighting continued until 1995 with the total number of deaths placed around 20,000 (Presidents Apologise Over Croatian War, 2003) and the total number of refugees and displaced persons placed around half a million (Frucht, 2005). Even though more than two decades have passed since the formal resolution of the

conflict, overt displays of prejudice and violence still occur, with the most recent high-profile case being the incident involving football fans at the 2022 World Cup (BBC, 2022).

Participants. Eligible participants had to identify as either ethnic Croats or Serbs, be nationals of, and live in Croatia or Serbia, and have no history of diagnosed mental health disorders or autism. Although Croatia and Serbia are highly ethnically homogeneous, in Croatia a small minority identify as ethnic Serbs (4.36%, Croatian Bureau of Statistics, 2011) and in Serbia a small minority identify as ethnic Croats (0.8%, Statistical Office of the Republic of Serbia, 2018). As I wanted to avoid recruiting ethnic Croats living in Serbia and/or having a Serbian nationality and vice versa, participants had to have a common identity across current place of residence, ethnicity, and nationality to be included in the experiment. A total of 371 participants were recruited through social media and completed the experiment in Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Twenty-five participants were excluded because they failed one or more of the three attention checks, leaving a final sample of 346 participants (152 men, 192 women, 2 preferred not to disclose; 161 Croatian; 185 Serbian; M_{AGE} = 25.02, SD_{AGE} = 7.22).

With 346 participants and alpha = .05, Experiment 4 had 80% power to detect a small effect size equivalent to d = 0.11 (Faul et al., 2009). However, since there is no established way to derive commonly reported effect sizes (e.g., partial eta squared, Cohen's d) from the output of linear mixed models (Judd et al., 2012) and since classical approaches to power and sensitivity analyses fail to account for the fact that linear mixed models capture multiple sources of random variation (Kumle et al., 2021), I also conducted post hoc sensitivity simulations (1000 iterations, Green & MacLeod, 2016) specific to each linear mixed model reported in the "Results" section. These sensitivity simulations revealed that the observed

estimates were, on average, 172.22% larger (median = 149.03%) than the minimal detectable estimate with 80% power (range = [11.61%, 473.04%][50]).

Contextualised perceptions of emotion-perception and empathic norms. After completing the consent form, participants were presented with pre-selected short stories – eight describing negative events experienced by their ethnic ingroup members, and eight describing matched events experienced by their ethnic outgroup members. The two sets of short stories were labelled as belonging to ingroup or outgroup members interchangeably across participants. Since both Croatian and Serbian language have grammatical gender and to prevent participants from assuming that any of the stories belong to their ethnic ingroup member living in an outgroup country (an ethnic Croat living in Serbia or an ethnic Serb living in Croatia), in addition to the country name and flag cue informing participants of the empathic targets' group category membership, each story was also randomly linked to a male first name, stereotypically perceived as ethnically Croatian or Serbian (see Appendix 3.1).

The short stories were written in first-person narrative where different empathic targets talked about a negative event they had directly experienced in the past. All stories described personal experiences that were unrelated to the empathic targets' ethnic identities or their experience of conflict (for example, in two stories, different empathic targets described their experiences of losing a pet or getting stuck in an elevator). Each story had a specific part clearly marked at or close to its end signifying the peak of the described emotional experience and participants were asked to focus on the marked section when giving responses. The final 16 short stories were chosen based on a pilot study. More details about the pilot study can be found in Appendix 3.2, while the full list of stories can be found

[50] Only one observed estimate fell below the minimal detectable estimate with 80% power, and this estimate will be labelled in the results section.

on Open Science Framework under Experiment 1 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d).

After reading each story, participants gave emotion perception, empathy and schadenfreude estimates from their own perspective (emotion perception: *How bad/good*[51] *do you think this person felt?*; empathy/schadenfreude: *How bad/good does what happened to this person make you feel?*, questions adapted from Cikara et al., 2014), from the perspective of an average ingroup and an average outgroup member (emotion perception: *How bad/good would an average [ingroup/outgroup member] think this person felt?*; empathy/schadenfreude: *How bad/good would what happened to this person make an average [ingroup/outgroup member] feel?*), with the order of questions counter-balanced across participants. Trial-by-trial empathy responses that participants gave from the perspective of an average ingroup and outgroup member constitute perceptions of ingroup and outgroup empathic norms[52], while the analogous estimates of emotion constitute perceptions of emotion-perception norms.

Individual-differences measures. Participants also completed a measure of prejudice, and ingroup and outgroup prejudice perceptions (Converse et al., 1980; Haddock et

[51] The positively worded emotion perception and schadenfreude scales gave results that were the exact inverse of results from negatively worded scales. Thus, for brevity, I only report results from negatively worded scales in this chapter. Of note, despite the correlation between the negatively worded and the positively worded scales (emotion perception: $r = -.65$, $p < .001$; empathy: $r = -.59$, $p < .001$), these scales were never intended to be combined into a single score as they have previously been shown to load on separate factors corresponding to empathy and schadenfreude respectively (see Hudson et al., 2019 for a factor analysis, also see Bruneau et al., 2015 & Cikara et al., 2014 for a similar approach). While negatively and positively worded emotion perception scales could arguably be combined, I decided not to do that to maintain consistency with empathy and schadenfreude scales.

[52] Please note that *perceptions of empathic norms* refer to how much feelings of outgroup members are perceived as shared in one's reference group, not to the process people themselves engage in to infer intergroup empathic norms (e.g., perspective taking).

al., 1993) as well as reporting their acontextualised perceptions of empathic norms and perceptions of ingroup and outgroup members' trait empathy[53].

Prejudice and prejudice perceptions. Prejudice and prejudice perceptions were assessed using adapted feeling thermometers (Converse et al., 1980; Haddock et al., 1993). Participants reported their own feelings as well as predicted feelings of average ingroup and outgroup members towards different groups, including the ingroup and outgroup. They gave their responses on 100-point slider scales, where 0 meant "I/They have very negative feelings towards [group]" depending on the perspective, and 100 meant "I/They have very positive feelings towards [group]". These measures were recoded so that higher scores indicated greater prejudice (more negative feelings) towards target groups.

Perceptions of trait empathy and acontextualised perceptions of empathic norms. In addition to assessing perceptions of empathic norms in the context of single empathic targets and their emotional experiences, I also assessed these perceptions in the abstract, in the absence of any specific targets or emotional context. Specifically, participants were asked how empathic they assumed ingroup and outgroup members to be towards ingroup and outgroup empathic targets, without any reference to individual targets or emotional circumstances (*Imagine something bad happened to someone who is Croatian/Serbian. How much empathy would you expect an average Croatian/Serbian person to feel in response?*). Additionally, participants were asked how empathic they perceived ingroup and outgroup members to be in general (*How empathic do you find an average Croatian/Serbian person?*). Prior to giving any responses, empathy was defined to participants as follows: "In

[53] In addition to the two measures described in the main text, participants completed a measure of dispositional empathic concern and perspective taking (Davis, 1983), ingroup identification (Doosje et al., 1995; Leach et al., 2008), and intergroup contact frequency (Barlow et al., 2012). For more details on these measures used in control analyses or included for a different purpose, see the Open Science Framework repository under Experiment 1 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d).

psychology, empathy refers to situations when a person sees someone expressing an emotion and starts to feel a similar emotion. For instance, the person sees someone crying and feels sad with them, or the person hear about a friend's promotion and feels excited with them." (Depow et al., 2021). All responses were given on 100-point slider scales, where 0 meant "They would feel no empathy/I don't see them as empathic at all" and 100 meant "They would feel total empathy/I see them as very empathic".

Analysis strategy. I analysed data from all experiments using linear mixed models implemented in R (version 4.0.4) package lme4 (version 1.1.27.1, Bates et al., 2014), and lmerTest (version 3.1.3, Kuznetsova et al., 2017)[54]. All pairwise comparisons used the Holm method for p-value correction from the package lsmeans (version 2.30.0, Lenth, 2016). Regarding the models' random structure, maximal models were always fitted first (Barr et al., 2013), and convergence and singularity warnings were resolved by simplifying the random structure using Principal Component Analysis to determine the most relevant random components (Bates et al., 2015). The random structure was always identified prior to adding fixed effects to the model (Meteyard & Davies, 2020). More information about specific analyses which yielded the results reported below are provided at the beginning of the relevant sections.

Results

Acontextualised perceptions of empathic norms. To assess whether participants expected intergroup empathy bias from average ingroup and outgroup members (research question a) and whether there was any own-group-favouring asymmetry in these expectations

[54] Given how variance is partitioned in linear mixed models (Rights & Sterba, 2019), there is no established way to derive commonly reported effect sizes (e.g., partial eta squared, statmerkur, 2019). Thus, we follow the recommendations of Pek and Flora (2018) and report unstandardised effect sizes (e.g., mean differences) for relevant comparisons.

(research question d), I regressed acontextualised perceptions of empathic norms onto fixed effects for empathiser (levels: average ingroup member, average outgroup member), empathic target (levels: own-group member, other-group member) and participants' ethnicity (levels: Croatian, Serbian) and included random intercepts for participants. Separately, I regressed perceptions of trait empathy onto fixed effects for empathiser (levels: average ingroup member, average outgroup member) and participants' ethnicity (levels: Croatian, Serbian) and included random intercepts for participants.

Although participants perceived ethnic ingroup and outgroup members as equally empathic overall (b = 1.04, 95% CI = [-0.72, 2.80], t(342) = 1.16, p = .247), they expected more intra- than intergroup empathy from both an average ingroup member (Croatian sample: b = 20.03, 95% CI = [15.58, 24.48], t(1026) = 11.90, p < .001; Serbian sample: b = 22.92, 95% CI = [18.75, 27.10], t(1026) = 14.51, p < .001[55]) and an average outgroup member (Croatian sample: b = 17.50, 95% CI = [13.05, 21.95], t(1026) = 10.39, p < .001; Serbian sample: b = 28.40, 95% CI = [24.23, 32.58], t(1026) = 17.98, p < .001, see Figure 3.1). The effect of empathic targets' group category membership on acontextualised perceptions of empathic norms was qualified by a two-way interaction between targets' group membership and empathiser (F(1, 1026) = 17.86, p < .001) as well as by a three-way interaction between targets' group membership, empathiser, and participants' ethnicity (F(1, 1026) = 6.03, p = .014). While Croatian participants did not display any asymmetry in their acontextualised perceptions of empathic norms, with average ingroup, and outgroup members expected to show equally strong empathic responses towards own-group members (b = -1.49, 95% CI = [-5.94, 2.96], t(1026) = -0.89, p = .752) as well as towards other-group members (b = 1.04, 95% CI = [-3.41, 5.50], t(1026) = 0.62, p = .752), I detected *empathic norm*

[55] Estimates are reported separately for each sample only for the models in which interactions with participants' ethnicity were significant.

perception asymmetry in the Serbian sample. Specifically, Serbian participants assumed that an average outgroup member would discriminate more in their empathy than an average ingroup member. This assumption stemmed from expectations that an average outgroup member would show lower other-group empathy than an average ingroup member (b = -4.44, 95% CI = [-0.27, -8.62], $t(1026) = -2.81$ $p = .010$), and not from expectations that they would show stronger own-group empathy than an average ingroup member (b = 1.04, 95% CI = [-3.14, 5.21], $t(1026) = 0.66$, $p = .511$). The discrepancy between Croatian and Serbian samples was unanticipated but might be associated with the fact that recruited Serbian participants had more positive feelings for their ethnic ingroup in comparison to recruited Croatian participants ($F(1, 342) = 11.60$, $p < .001$).

Contextualised perceptions of emotion-perception and empathic norms. Next, I assessed whether participants' contextualised responses also reflected an assumption that average ingroup and outgroup members would show intergroup bias in empathy and emotion perception (research question a). Furthermore, I compared participants' own intergroup emotion perception and empathy biases to biases they assumed average ingroup and outgroup members would show (research question b). Specifically, I regressed combined emotion perception and empathy responses onto fixed effects for norm type (levels: emotion perception, empathy), empathiser (levels: self, average ingroup member, average outgroup member), target (levels: own-group member, other-group member) and participants' ethnicity (levels: Croatian, Serbian). I included random intercepts for participants and stories and a correlated random slope for empathic target within participant.

When participants reported perceptions of emotion-perception and empathic norms in relation to specific targets and their emotional circumstances, I observed a significant three-way interaction between empathic target's group membership, empathiser, and norm type as expected ($F(2, 32489) = 8.67$, $p < .001$). The general pattern of responses remained the same

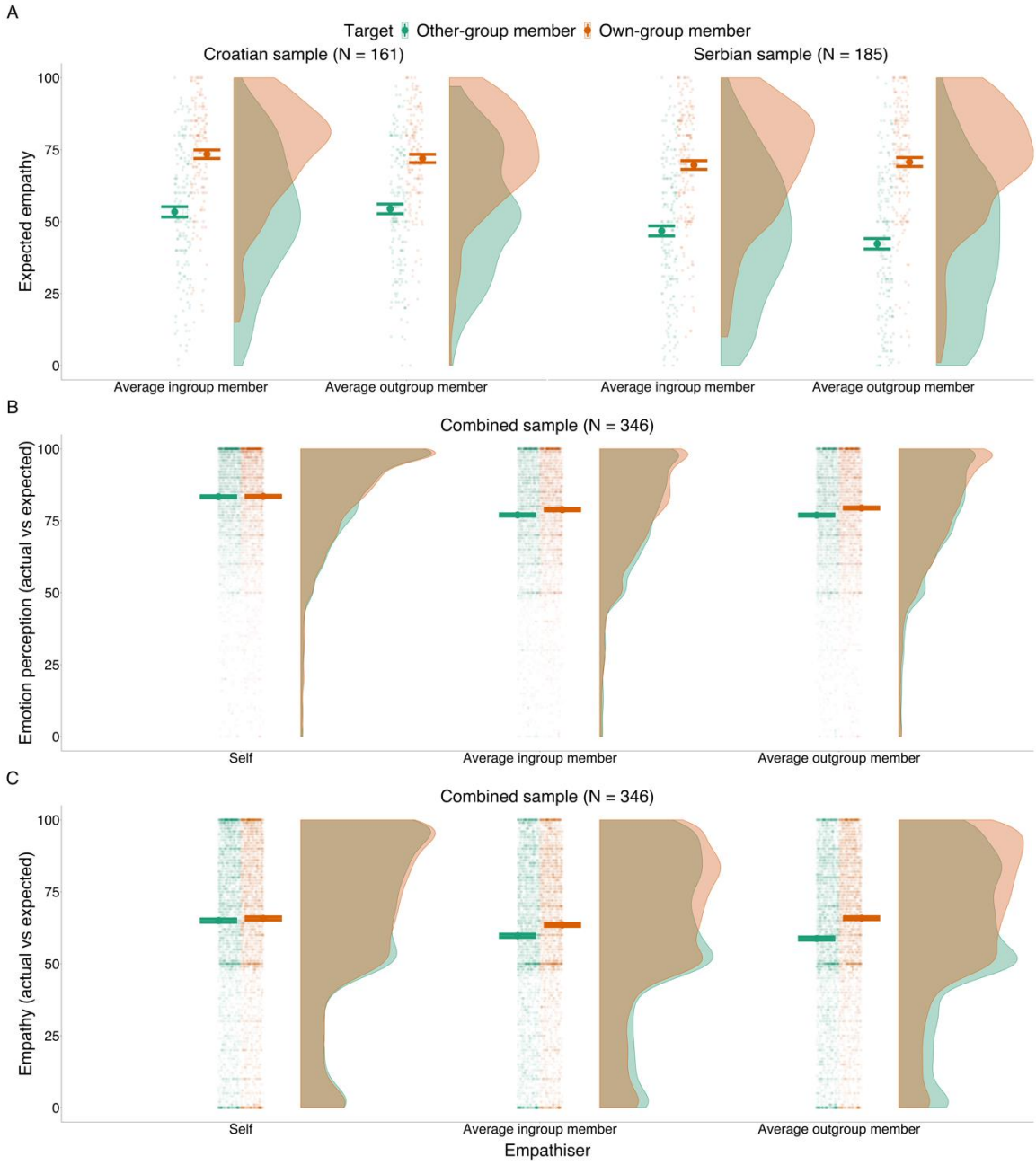
as in the case of acontextualised perceptions of empathic norms – participants again expected greater own-group than other-group empathy from both an average ingroup member (b = 3.82, 95% CI = [2.34, 5.31], z = 7.55, p < .001) and an average outgroup member (b = 6.88, 95% CI = [5.40, 8.37], z = 13.59, p < .001). In comparison to intergroup empathy bias, participants reported expecting a smaller intergroup emotion perception bias such that both an average ingroup (b = 1.93, 95% CI = [0.44, 3.41], z = 3.80, p < .001) and an average outgroup member (b = 2.31, 95% CI = [0.83, 3.80], z = 4.57, p < .001) were expected to perceive own-group targets as feeling more negative than other-group targets about their emotional circumstances. By contrast, empathic targets' group membership did not in fact influence participants' own emotion perception (b = 0.25, 95% CI = [-1.23, 1.74], z = 0.50, p > .999) or empathy responses (b = 0.80, 95% CI = [-0.68, 2.29], z = 1.59, p = .337, see Figure 3.1).

Next, to directly test whether participants assumed that outgroup members discriminate more than ingroup members in their empathy (research question d), but not emotion perception, I fitted the previous model with participants' own emotion perception and empathy responses excluded. If there is asymmetry in the perceptions of empathic, but not emotion-perception norms, I would expect to observe a three-way interaction between norm type (levels: emotion perception, empathy), empathiser (levels: average ingroup member, average outgroup member) and target (levels: own-group member, other-group member), with planned comparisons illuminating the source of asymmetry. After excluding participants' responses from their own perspective, I observed this significant three-way interaction (F(1, 21425) = 7.60, p = .006). Pairwise comparisons revealed that participants expected greater intergroup empathy bias from an average outgroup member than from an average ingroup member, such that an average outgroup member was assumed to empathise more with own-group targets in comparison to an average ingroup member (b = 2.21, 95% CI

= [0.93, 3.49], z = 4.56, p < .001), while the expected difference between average ingroup and outgroup members in their empathy for other-group targets was not statistically significant (b = -0.85, 95% CI = [-2.13, 0.43], z = -1.75, p = .081). In contrast to the domain of empathy, no asymmetry was found in the domain of emotion perception. Specifically, no differences were expected between an average ingroup, and an outgroup member in how they perceive own-group targets' (b = 0.42, 95% CI = [-0.86, 1.70], z = 0.87, p = .764) and other-group targets' emotions (b = 0.04, 95% CI = [-1.24, 1.32], z = 0.09, p = .931).

Figure 3.1

Perceptions of emotion-perception and empathic norms (Experiment 4)



Notes. (A) Acontextualised perceptions of empathic norms reported in the abstract, in the absence of any information about specific targets and emotional experiences. The higher the values on the y-axis, the more empathy participants assumed that average ingroup and outgroup members would show for an abstract empathic target. (B) Contextualised emotion perception and perceptions of emotion-perception norms. The higher the values on the y-axis, the more negative participants thought the targets felt (self) or the more negative they assumed average ingroup and outgroup members would think the targets felt, depending on the perspective (average ingroup/outgroup member). (C) Contextualised empathy perceptions of empathic norms. The higher

the values on the y-axis the more negative participants felt in response to the targets' experiences (self) or the more negative they assumed average ingroup and outgroup members would feel in response to the targets' experiences, depending on the perspective (average ingroup/outgroup member). Lighter dots in the background represent participants' individual responses, while solid dots and error bars in the foreground represent means and standard errors. Results are presented separately for each sample only where the interaction with participants' ethnicity was significant.

Accuracy of emotion-perception and empathic norm perceptions. Next, I compared the intergroup emotion perception and empathy biases participants expected with the actual intergroup emotion perception and empathy biases respective ingroup and outgroup members reported (research question c). I derived estimates of intergroup emotion perception and empathy biases from first-person perspective responses and these estimates were used as ground truth for comparison with intergroup emotion perception and empathy biases participants expected. For example, Croatian participants' prediction of how much an average Serbian person would discriminate in their empathy was compared to how much sampled Serbian participants actually discriminated in their empathy. When comparing participants' predictions of how ingroup members would respond with how actual ingroup members responded, participants' own responses were always excluded from the calculation of the ingroup estimate. Finally, even though I made sure that crucial demographic characteristics (e.g., ethnic identity, gender) were represented to an equal degree, the samples are not representative of the ethnicities in question. Thus, I make no assumption that the ground truth estimates I collected were fully representative of the relevant populations but believe that they still provide a useful approximation.

As expected, both Croatian and Serbian participants overestimated the intergroup empathy bias of both ingroup (Croatian participants: $b = 2.77$, 95% CI = [0.22, 5.76], $z =$

2.72, $p = .039$[56]; Serbian participants: $b = 3.27$, 95% CI = [0.48, 6.06], $z = 3.44$, $p = .005$) and outgroup members (Croatian participants: $b = 4.20$, 95% CI = [1.06, 7.33], $z = 3.93$, $p < .001$; Serbian participants: $b = 8.38$, 95% CI = [5.24, 11.51], $z = 7.84$, $p < .001$). Correspondingly, participants overestimated how much ingroup ($b = 1.67$, 95% CI = [0.35, 2.99], $z = 3.03$, $p = .005$) and outgroup members ($b = 2.27$, 95% CI = [0.94, 3.59], $z = 4.10$, $p < .001$) perceived own-group members as feeling worse than other-group members in response to equivalent emotional experiences.

Comparing acontextualised and contextualised perceptions of empathic norms. To test the debiasing potential of providing contextual information about specific targets and their emotional circumstances (research question e), I regressed combined acontextualised and contextualised perceptions of empathic norms onto fixed effects for norm perception (levels: acontextualised, contextualised), empathiser (levels: average ingroup member, average outgroup member), target (levels: own-group member, other-group member) and participants' ethnicity (levels: Croatian, Serbian). I included random intercepts for participants and a correlated random slope for norm perception within participant.

As could already be seen by comparing Figures 3.1A and 3.1C, contextualisation reduced group membership's effect on perceptions of empathic norms. Specifically, participants expected significantly smaller intergroup empathy bias when context containing information about specific targets and their emotional experiences was introduced ($b = 5.43$, 95% CI = [4.70, 6.17], $z = 14.59$, $p < .001$) than when contextual information was absent ($b = 22.43$, 95% CI = [20.40, 24.51], $z = 21.23$, $p < .001$).

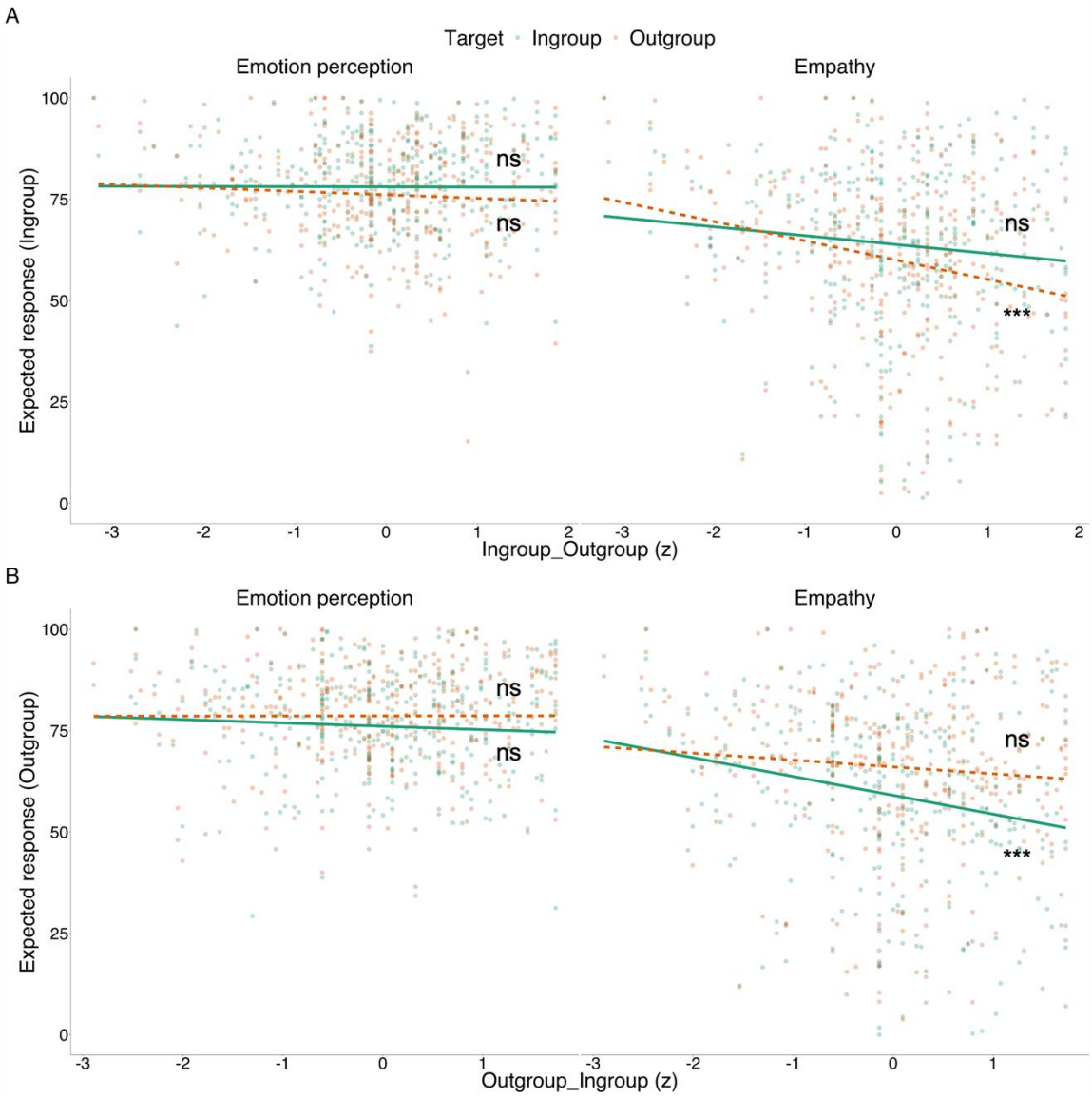
[56] In my sensitivity analyses, I found that this estimate fell below the minimal detectable estimate with 80% power. When the dependent variable was standardised, the minimal detectable estimate with 80% power was 0.154, while the observed estimate was 0.146.

Relationship between perceptions of group relations and perceptions of emotion-perception and empathic norms. Finally, to test the connection between perceptions of group relations, captured in perceptions of ingroup and outgroup prejudices, and perceptions of emotion-perception and empathic norms (research question f), I first fitted linear mixed models separately for the average ingroup and the average outgroup empathiser and included target (levels: own-group member, other-group member) and norm type (levels: emotion perception, empathy) as fixed factors of interest in the baseline model. I then fitted models which included additional fixed effects – prejudice perceptions in relevant directions (e.g., if perceptions of ingroup empathic norms were the dependent variable, relevant prejudice perceptions were people's beliefs about ingroup feelings towards the ingroup and the outgroup) and compared the new models to the baseline model using information criteria (AIC, BIC). I report only results from the best performing models, while details about model selection can be found in Appendix 3.3.

In terms of results, models which included relevant intergroup prejudice perceptions – people's beliefs about how ingroup members felt towards the outgroup and how outgroup members felt towards the ingroup – significantly outperformed all other models. These models revealed that the more negative feelings participants thought the ingroup had towards the outgroup or vice versa, the less empathy (ingroup empathic norms: b = -4.55, SE = 1.17, t = -3.89, $p < .001$; outgroup empathic norms: b = -4.02, SE = 1.18, t = -3.42, $p < .001$) they expected in the same direction (i.e., ingroup towards outgroup or outgroup towards ingroup respectively, see Figure 3.2). Importantly, these same-direction findings were specific to empathy, and intergroup prejudice perceptions did not reliably predict relevant perceptions of emotion-perception norms (ingroup emotion perception norms: b = -0.71, SE = 1.17, t = -0.61, p = .550; outgroup emotion perception norms: b = 0.39, SE = 1.18, t = 0.34, p = .740).

Figure 3.2

Perceived emotion-perception and empathic norms in relation to intergroup prejudice perceptions (Experiment 4)



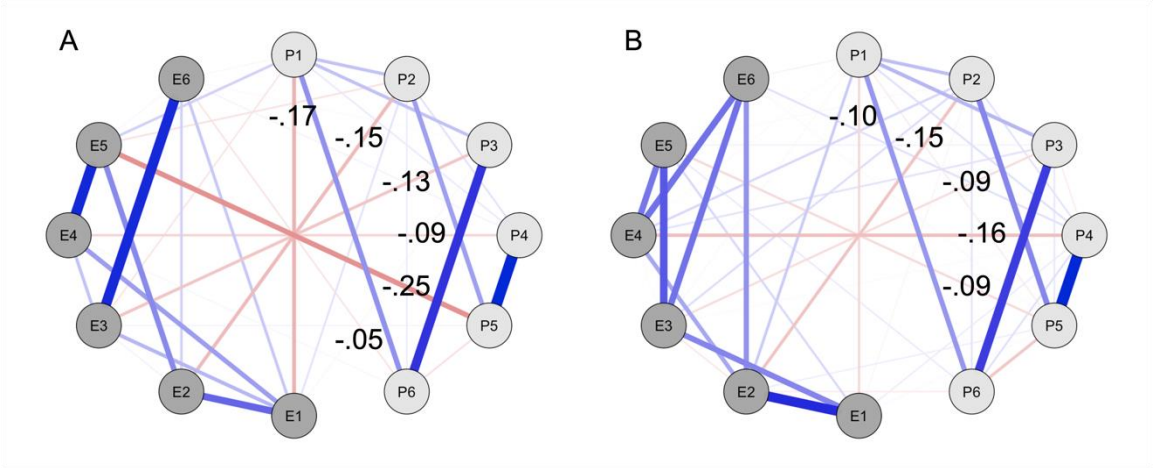
Notes. The prejudice perception scales were reversed and standardised such that higher values indicate that (A) participants believed ingroup members harboured more negative feelings towards the outgroup in general, and (B) outgroup members harboured more negative feelings towards the ingroup in general (r = .74, p < .001). Due to the high correlation between the two prejudice perceptions, the irrelevant, but correlated perception was always controlled for to isolate the directional prediction and prevent spillover effects. ns = Slope not significant. *** Slope significant at p < .001.

To confirm that perceptions of ingroup and outgroup prejudices uniquely predicted same-direction perceptions of empathic norms, I also fitted undirected partial correlation

networks including own prejudices, and prejudice perceptions on one side and acontextualised perceptions of trait empathy and empathic norms (see Figure 3.3A) or averaged contextualised empathy and perceptions of empathic norms (see Figure 3.3B) on the other side. I examined the strength of inter-domain connections (edges), partial correlations between prejudice-related and empathy-related nodes, while I did not interpret within-domain connections (e.g., connections between two empathy-related or two prejudice-related nodes). Importantly, if an edge connects two nodes, that means that these nodes' association cannot be explained by other nodes in the network (Epskamp & Fried, 2018). To avoid over-interpreting potentially spurious connections, I used LASSO regularisation, and the gamma tuning parameter was chosen to minimise the extended Bayesian information criterion (Epskamp & Fried, 2018). As network analyses are demanding in terms of sample size (Epskamp et al., 2018), I only estimated these networks using data from all participants, and not in each sample separately. Finally, I assessed the accuracy of estimated network connections following Epskamp, Borsboom, and Fried's (2018) recommendations (see Appendix 3.4). Regardless of whether acontextualised (Figure 3.3A) or contextualised (Figure 3.3B) responses were included as nodes, the network analyses confirmed that the most prominent inter-domain connections were between same-direction nodes – for instance, the more negative feelings ingroup members were thought to have towards the outgroup in general, the less they were expected to empathise with specific outgroup targets (P4-E4 edge, see Figure 3.3).

Figure 3.3

Empathy and perceptions of empathic norms in relation to prejudices and prejudice perceptions (Experiment 4)



Notes. Regularised partial correlation networks based on (A) participants' acontextualised perceptions of ingroup and outgroup members' trait empathy and empathic norms, and (B) participants' contextualised empathy and perceptions of empathic norms. E1 = Perceptions of ingroup members as empathic (A) or one's own empathy for ingroup targets (B). E2 = Perceptions of outgroup members as empathic (A) or one's own empathy for outgroup targets (B). E3 = Expected empathy of ingroup members for ingroup targets. E4 = Expected empathy of ingroup members for outgroup targets. E5 = Expected empathy of outgroup members for ingroup targets. E6 = Expected empathy of outgroup members for outgroup targets. P1 = One's own prejudice towards ingroup targets. P2 = One's own prejudice towards outgroup targets. P3 = Expected prejudice of ingroup members towards ingroup targets. P4 = Expected prejudice of ingroup members towards outgroup targets. P5 = Expected prejudice of outgroup members towards ingroup targets. P6 = Expected prejudice of outgroup members towards outgroup targets. Displayed values represent strengths of significant edges.

Discussion

In accordance with the first hypothesis, results from Experiment 4 showed that perceptions of emotion-perception and especially empathic norms tend to be overly negative and inaccurate in the post-conflict context of Croatia and Serbia, with participants assuming that ingroup and outgroup members would be more affected than themselves by empathic targets' group category membership in the way they perceive and empathise with emotions (research question b) and assuming more bias in emotion perception and empathy than was

shown by actual ingroup and outgroup members (research question c). In line with the second hypothesis, Experiment 4 provided initial evidence of *empathic norm perception asymmetry*, wherein outgroup members were assumed to discriminate more than ingroup members in their empathy, but not in the way they perceive emotions. In other words, participants assumed outgroup members' intergroup empathy bias to be greater than ingroup members' while I found no differences in participants' assumptions of outgroup and ingroup members' intergroup emotion perception bias. While there is evidence that equal empathy for everyone regardless of their group category membership is perceived as more morally right than biased empathy for close others (Fowler et al., 2021), I have no reason to assume that impartial emotion-perception judgments, akin to the ones participants provided in Experiment 4, have similar connotations of morality. Thus, it is possible that empathic norm perception asymmetry reflects people's general tendency to see ingroup members in a more positive light, including in terms of morality (Enock et al., 2021; Liberman et al., 2018).

Furthermore, in line with the third hypothesis, when participants reported their perceptions of empathic norms in relation to specific empathic targets and their emotional experiences, they expected a smaller intergroup empathy bias from ingroup and outgroup members than when they reported their perceptions of empathic norms as abstract assumptions, in the absence of specific targets and emotional experiences. It is worth noting that the narrative context presented to participants in Experiment 4 contained no information relating to empathic targets' social identities or experience of conflict, and it is therefore possible that the context facilitated construal of empathic targets beyond their social identities, encouraging perspective taking and empathy and reducing bias in perception of respective norms.

Finally, in accordance with the final exploratory hypothesis, perceptions of intra- and intergroup empathic norms were reliably predicted by corresponding perceptions of intra- and

intergroup relations' valence. This extends previous findings showing that one's own empathy is enhanced when empathic targets are liked (e.g., Marsh, 2022) by documenting that subjective perceptions of empathic norms in intergroup environments are also enhanced when the relevant within- and between-group relations are believed to be positive.

Experiment 5

In Experiment 5, I addressed the same set of research questions in the political intergroup context surrounding the UK's departure from the European Union ("Brexit") with the aim of establishing whether the results observed in Experiment 4 generalise across different intergroup contexts. I also adapted the research design to investigate whether participants' own emotion perception and empathy as well as their perceptions of respective norms vary as a function of how much empathic targets' emotional experiences make their social identities salient (research question e). Specifically, while Experiment 4 showed that being introduced to specific empathic targets and their emotional experiences reduced assumptions of intergroup bias, in Experiment 5, I tested whether these debiasing effects are reserved for situations where contextual information draws attention towards each target's unique emotional circumstances and away from their social identity, thus allowing for empathic targets to be seen as "individuals" rather than "group members". In other words, I hypothesised that the availability of contextual information about empathic targets would not similarly reduce assumptions of others' intergroup bias when the contextual information increased the salience of empathic targets' social (i.e., political) identity.

Method

Intergroup context. In a referendum held on June 23rd, 2016, a slim majority in the United Kingdom voted in favour of the country withdrawing from the European Union (52%), and four years later, on December 31st, 2020, this withdrawal process – commonly

referred to as “Brexit” – was completed. Beyond its geopolitical significance, Brexit had sweeping consequences for British citizens’ sense of political identity, with Brexit identities – “Leave(rs)” and “Remain(ers)” – replacing more traditional political party allegiances and continuing to generate intense tension long after the results of the referendum were known (Hobolt et al., 2021).

Participants. Eligible participants had to hold only UK citizenship and reside in the UK, have English as their first language and have no history of diagnosed mental health disorders or autism. A total of 202 participants were recruited through Prolific (www.prolific.co) and redirected to the experiment hosted by Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Seven participants were excluded because they failed two or more of the six attention checks, leaving a final sample of 195 participants (98 men, 97 women; 94 Leave supporters, 101 Remain supporters; M_{AGE} = 37.25, SD_{AGE} = 10.93). If participants failed one attention check, only the trial where they made the mistake was excluded ($N_{EXCLUDED}$ = 16). With 195 participants and alpha = .05, Experiment 5 had 80% power to detect an effect equivalent to d = 0.15 (Faul et al., 2009). Post hoc sensitivity simulations for fitted linear mixed models (1000 iterations, Green & MacLeod, 2016) also revealed that my observed estimates were, on average, 244.92% larger (median = 207.59%) than the minimal detectable estimate with 80% power (range = [19.03%, 661.18%]).

Contextualised perceptions of emotion-perception and empathic norms. Contextualised perceptions of emotion-perception and empathic norms were assessed in the same way as in Experiment 4. Participants were presented with pre-selected short stories – eight describing negative events experienced by members of their political ingroup, and eight describing matched events experienced by members of their political outgroup – and after reading each story, participants provided emotion perception, empathy, schadenfreude, and

compassion responses[57] from their own perspective as well as from the perspective of average ingroup and outgroup members. Importantly, whereas all stories used in Experiment 4 described experiences unrelated to the empathic targets' ethnic identities, to test the new hypothesis, half of the stories in Experiment 5 described experiences unrelated to the empathic targets' political identities (for example, in two of the stories, different targets described their experiences of losing a wallet or having to cancel a family trip due to personal illness), while the other half described experiences that played a role in shaping each target's political identity (for example, in two of the stories, different targets described their experiences of being unable to participate in the EU exchange programme or losing a job in the tourism industry due to uncertainty surrounding Brexit, both of which played a role in these targets deciding to vote Remain). I use the terms *group-irrelevant* and *group-relevant* experiences to refer to these two types of emotional experiences in my presentation of results. Furthermore, only experiences irrelevant to the targets' social identities were counter-balanced across participants, whereas experiences relevant to their social identities were always presented with the same Leave or Remain label (for example, the target who had a negative encounter with someone who turned out to be an undocumented immigrant was always a Leave supporter, while the target who suffered an injury could be either a Leave or a Remain supporter across participants). This was done to avoid combinations of experiences and identities being perceived as incongruent and unconvincing. Finally, similar to Experiment 4, the 16 stories were chosen based on a pilot study. More details about the pilot study can be found in Appendix 3.2, while the full list of stories can be found on Open

[57] As the positively worded emotion perception and schadenfreude scales again gave results that were the exact opposite of results from negatively worded scales, I only report results from negatively worded scales in this chapter. Despite their correlation (emotion perception: $r = -.57$, $p < .001$; empathy: $r = -.56$, $p < .001$), these scales were again not combined into a single score for the same reasons as in Experiment 4. Similarly, as participants' compassion responses were identical to their empathy responses, I only focus on empathy in this chapter.

Science Framework under Experiment 2 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d).

Individual-differences measures. Participants completed the same measure of prejudice and ingroup and outgroup prejudice perceptions as in Experiment 4 (Converse et al., 1980; Haddock et al., 1993). Additionally, participants were asked about their original Brexit referendum vote and current Brexit attitude (Zmigrod et al., 2018)[58] and completed a measure of Brexit attitude strength (Pomerantz et al., 1995). Finally, as in Experiment 4, participants reported their acontextualised perceptions of trait empathy and empathic norms by rating how empathic they perceived political ingroup and outgroup members to be in general, and towards unspecified ingroup and outgroup targets[59].

Results

Acontextualised perceptions of empathic norms. I followed the same analysis strategy as in Experiment 4. First, to address research questions a and d in the domain of acontextualised responses, I regressed acontextualised perceptions of empathic norms onto fixed effects for empathiser (levels: average ingroup member, average outgroup member), target (levels: own-group member, other-group member) and participants' Brexit identity (levels: Leave, Remain) and included random intercepts for participants. Separately, I regressed perceptions of trait empathy onto fixed effects for empathiser (levels: average

[58] I classified participants as Leave or Remain supporters based on their current Brexit attitude. However, all reported results remained the same even when I excluded participants who changed their Brexit attitude since the referendum (e.g., participants who originally voted Leave, but changed their attitude to Remain and vice versa).

[59] In this chapter, I focus on participants' acontextualised perceptions of trait empathy and empathic norms as well as on their prejudices and prejudice perceptions, but more details on other measures included for control analyses and for a different purpose can be found in the Open Science Framework repository under Experiment 2 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d).

ingroup member, average outgroup member) and participants' Brexit identity (levels: Leave, Remain) and included random intercepts for participants.

Participants again expected more intra- than intergroup empathy from both an average ingroup member (b = 20.31, 95% CI = [15.32, 25.30], t(579) = 10.78, p < .001) and an average outgroup member (b = 26.38, 95% CI = [21.39, 31.36], t(579) = 14.00, p < .001, see Figure 3.4), but in contrast to Experiment 4, they also viewed political ingroup members as more empathic in general (Leave supporters: b = 11.20, 95% CI = [5.05, 17.30], t(193) = 3.60, p < .001; Remain supporters: b = 21.80 95% CI = [15.93, 27.70], t(193) = 7.30, p < .001[60]).

Empathic norm perception asymmetry was evidenced by a significant two-way interaction between empathic targets' group membership and empathiser (F(1, 579) = 5.18, p = .023). As in Experiment 4, empathic norm perception asymmetry was a product of expectations that an average outgroup member would show less other-group empathy than would an average ingroup member (b = -8.76, 95% CI = [-3.77, -13.75], t(579) = 4.65, p < .001), and not of expectations that they would show more own-group empathy than would an average ingroup member (b = -2.70, 95% CI = [-7.68, 2.29], t(579) = 1.43, p = .153).

Contextualised perceptions of emotion-perception and empathic norms. To address research questions a and b in the domain of contextualised responses, in two models zooming in on emotion perception and empathy, I regressed respective responses onto fixed effects for empathiser (levels: self, average ingroup member, average outgroup member), target (levels: own-group member, other-group member), experience type (levels: group-irrelevant, group-relevant), and participants' Brexit identity (levels: Leave, Remain). I

[60] Estimates are reported separately for each sample only for the models in which interactions with participants' Brexit identity were significant.

included random intercepts for participants and stories and correlated random slopes for target and experience type within participant.

When contextual information about empathic targets and their emotional circumstances was provided, the general pattern of empathic norm perceptions remained the same – participants still expected greater own-group than other-group empathy from both average ingroup and outgroup members. However, the expected intergroup empathy bias was substantially larger for empathic targets' group-relevant experiences (average ingroup member: b = 18.52, 95% CI = [15.13, 21.91], z = 18.40, p < .001; average outgroup member: b = 22.29, 95% CI = [18.90, 25.68], z = 22.15, p < .001) than for their group-irrelevant experiences (average ingroup member: b = 3.80, 95% CI = [0.43, 7.18], z = 3.79, p = .005; average outgroup member: b = 3.44, 95% CI = [0.07, 6.82], z = 3.44, p = .017). By contrast, participants did not show intergroup empathy bias when empathic targets' emotional experiences were unrelated to their Brexit identity (b = 1.05, 95% CI = [-2.32, 4.43], z = 1.05, p > .999), but showed a relatively small intergroup empathy bias when empathic targets' emotional experiences were group-relevant (b = 7.37, 95% CI = [3.98, 10.76], z = 7.32, p < .001). However, even for group-relevant experiences, group membership's effect on participants' own empathy was significantly weaker than its effect on perceptions of empathic norms.

When reporting perceived emotion-perception norms, Leave supporters did not expect any intergroup bias in how targets' emotions stemming from their group-irrelevant experiences would be perceived (average ingroup member: b = 1.82, 95% CI = [-2.24, 5.89], z = 1.51, p > .999; average outgroup member: b = 1.98, 95% CI = [-2.09, 6.05], z = 1.64, p > .999). However, Leave supporters still assumed that average ingroup and outgroup members would perceive own-group targets as feeling more negatively in response to their group-relevant experiences in comparison to other-group targets (average ingroup member: b =

15.23, 95% CI = [11.14, 19.32], z = 12.55, p < .001; average outgroup member: b = 11.09, 95% CI = [7.01, 15.18], z = 9.14, p < .001). Similarly, Remain supporters reported assuming no intergroup bias in how targets' emotions stemming from their group-irrelevant experiences would be perceived (average ingroup member: b = 0.73, 95% CI = [3.41, 4.87], z = 0.59, p > .999; average outgroup member: b = 3.72, 95% CI = [-0.42, 7.86], z = 3.03, p = .052), but continued to assume that own-group targets' emotions in response to group-relevant experiences would be perceived as more negative than those of other-group targets by both average ingroup (b = 12.35, 95% CI = [8.19, 16.51], z = 10.00, p < .001) and outgroup members (b = 21.93, 95% CI = [17.77, 26.09], z = 17.76, p < .001). In fact, none of the conditions revealed evidence of an intergroup bias in how participants perceived targets' emotions (all ps > .999), except for Leave supporters who showed a tendency to rate own-group targets' emotions triggered by group-relevant experiences as more negative than other-group targets' emotions (b = 4.81, 95% CI = [0.72, 8.89], z = 3.96, p = .002). However, even in this case, empathic targets' group category membership had a stronger effect on participants' perceptions of emotion-perception norms than on their own emotion perception responses.

Next, using the same approach as in Experiment 4, to directly test whether norm perception asymmetry occurred in participants' perceptions of empathic norms, but not emotion-perception norms (research question d), I refitted the models after excluding responses from participants' own perspective. If there is asymmetry in the perceptions of empathic norms regardless of the group relevance of the experience, I would expect to observe a two-way interaction between empathiser (levels: average ingroup member, average outgroup member) and target (levels: own-group member, other-group member). If the empathic norm perception asymmetry is dependent on the group relevance of the experience, I would expect to observe a three-way interaction between empathiser (levels: average

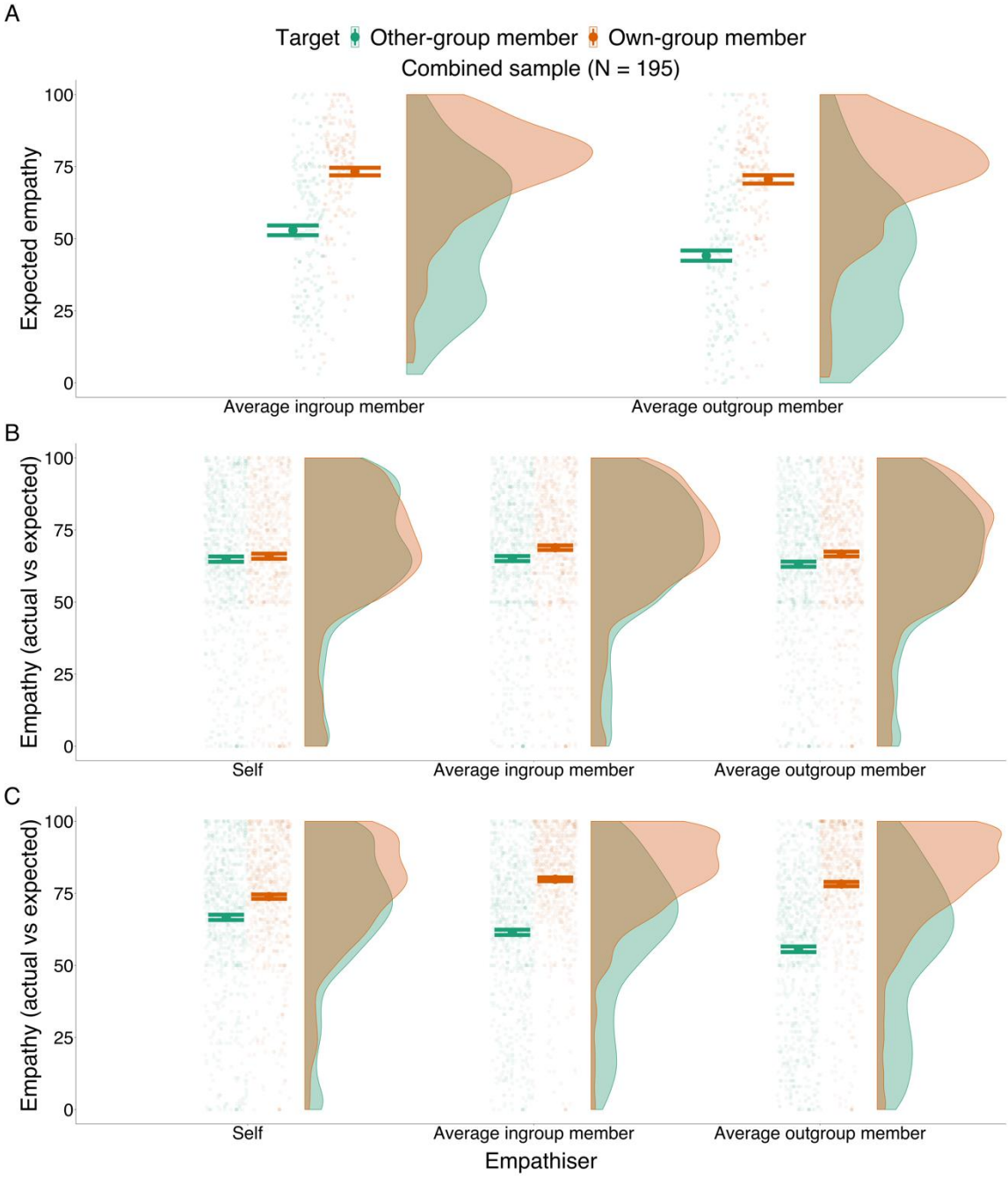
ingroup member, average outgroup member), target (levels: own-group member, other-group member) and experience type (levels: group-irrelevant, group-relevant) with planned comparisons clarifying the source of asymmetry. Finally, since the norm perception asymmetry in Experiment 4 was specific to perceptions of empathic norms, I did not expect to observe the described interactions with perceptions of emotion-perception norms as the dependent variable.

After excluding participants' own empathic responses, I observed a three-way interaction between empathiser, target, and experience type ($F(1, 5600) = 5.76$, $p = .016$). Pairwise comparisons revealed that while there was no norm perception asymmetry in participants' expectations in the domain of group-irrelevant experiences, participants assumed that an average outgroup member would be more biased in their empathy than an average ingroup member in the domain of group-relevant experiences. In other words, there were no significant differences between how much an average ingroup and an average outgroup member were assumed to empathise with own-group targets' group-irrelevant experiences ($b = 2.18$, 95% CI = [-0.50, 4.87], $z = 2.54$, $p = .078$), nor were there significant differences in how much empathy the two empathisers were expected to show for other-group targets' group-irrelevant experiences ($b = 1.83$, 95% CI = [-0.86, 4.51], $z = 2.13$, $p = .168$). However, the identified empathic norm perception asymmetry in the domain of group-relevant experiences was a consequence of expectations that an average outgroup member would show lower empathy for Brexit-related experiences of the other group than would an average ingroup member ($b = -5.57$, 95% CI = [-8.27, -2.87], $z = 6.45$, $p < .001$), and not a consequence of expectations that they would show higher empathy for Brexit-related experiences of the own group than would an average ingroup member ($b = -1.79$, 95% CI = [-4.49, 0.91], $z = 2.07$, $p = .168$).

Directly comparing perceptions of empathic norms in relation to group-irrelevant and group-relevant emotional experiences provides another approach to understanding the observed empathic norm perception asymmetry. Specifically, I found that participants expected an average ingroup member to empathise more with ingroup-relevant than with ingroup-irrelevant emotional experiences (b = 11.04, 95% CI = [4.22, 17.86], z = 5.06, p < .001), but participants did not expect an average ingroup member to feel differently about outgroup-relevant than outgroup-irrelevant emotional experiences (b = -3.67, 95% CI = [-10.49, 3.16], z = -1.68, p = .279). This expectation matched participants' actual empathy responses – their intergroup empathy bias in the domain of group-relevant emotional experiences was always a consequence of higher empathy for ingroup-relevant than for ingroup-irrelevant emotional experiences (b = 7.97, 95% CI = [0.14, 15.80], z = 3.43, p = .017) rather than lower empathy for outgroup-relevant than for outgroup-irrelevant emotional experiences (b = 1.66, 95% CI = [-6.17, 9.48], z = 0.71, p > .999). However, when reporting perceptions of outgroup empathic norms, participants expected not only higher empathy for outgroup-relevant emotional experiences (b = 11.43, 95% CI = [4.61, 18.25], z = 5.23, p < .001), but also lower empathy for ingroup-relevant emotional experiences (b = -7.41, 95% CI = [-14.23, -0.59], z = -3.40, p = .007) than for their group-irrelevant counterparts.

Figure 3.4

Perceptions of empathic norms (Experiment 5)



Notes. (A) Acontextualised perceptions of empathic norms reported in the abstract, in the absence of any contextual information about specific targets and emotional experiences. The higher the values on the y-axis, the more empathy participants assumed average ingroup and outgroup empathisers would show for an abstract empathic target. (B) Contextualised empathy and perceptions of empathic norms reported in relation to the empathic targets' group-irrelevant emotional experiences. (C) Contextualised empathy and perceptions of empathic norms reported in relation to the empathic targets' group-relevant emotional experiences. (B-C) The

higher the values on the y-axis the more negative participants felt in response to the empathic targets' experiences (self) or the more negative they assumed average ingroup and outgroup members would feel in response to the empathic targets' experiences, depending on the perspective (average ingroup/outgroup member). (A-C) Lighter dots in the background represent participants' individual responses, while solid dots and error bars in the foreground represent means and standard errors. Since I never observed significant interactions with participants' Brexit identities, all subplots combine the two political samples.

Finally, after excluding participants' own emotion-perception responses, contrary to my expectations, I observed a two-way interaction between empathiser and target (F(1, 5600.3) = 6.94, p = .008). This interaction revealed that participants assumed that an average outgroup member would be more biased by empathic targets' group category membership in the way they perceive emotions than an average ingroup member, regardless of whether the empathic targets' emotional experiences were group-relevant or group-irrelevant. This norm perception asymmetry was a consequence of participants' expectations that an average outgroup member would perceive other-group empathic targets' emotions as less negative than an average ingroup member (b = -3.93, 95% CI = [-5.44, -2.41], z = 6.82, p < .001).

Accuracy of emotion-perception and empathic norm perceptions. As in Experiment 4, I tested the accuracy of participants' norm perceptions by comparing the intergroup emotion perception and empathy biases participants expected with the intergroup emotion perception and empathy biases real ingroup and outgroup members in fact showed (research question c). For group-irrelevant emotional experiences, even though participants expected ingroup and outgroup members to be somewhat more biased in their empathy than they themselves were (see Figure 3.4), this expected bias did not significantly differ from the actual difference targets' group category membership made in ingroup (Leave supporters: b = 2.67, 95% CI = [-3.07, 8.40], t(2135) = 1.37, p > .999; Remain supporters: b = 2.83, 95% CI = [-2.70, 8.36], t(2135) = 1.51, p > .999) and outgroup members' empathy (Leave supporters: b = 0.98, 95% CI = [-5.30, 7.26], t(1298) = 0.46, p > .999; Remain supporters: b = 3.91, 95% CI = [-2.37, 10.19], t(1298) = 1.83, p = .943). For group-relevant emotional experiences,

however, participants' norm perceptions substantially exaggerated the intergroup empathy bias of both ingroup (Leave supporters: b = 9.76, 95% CI = [4.02, 15.49], t(2135) = 5.00, p < .001; Remain supporters: b = 12.52, 95% CI = [6.98, 18.05], t(2135) = 6.65, p < .001) and outgroup members (Leave supporters: b = 9.99, 95% CI = [3.72, 16.27], t(1298) = 4.68, p < .001; Remain supporters: b = 19.84, 95% CI = [13.56, 26.12], t(1298) = 9.29, p < .001).

In addition to overestimating ingroup and outgroup members' intergroup empathy bias, participants were also inaccurate in their assumptions of how biased ingroup and outgroup members would be in their judgments of empathic targets' emotions. Specifically, participants overestimated how much ingroup (Leave supporters: b = 10.34, 95% CI = [5.36, 15.31], t(2135) = 6.11, p < .001; Remain supporters: b = 10.67, 95% CI = [5.87, 15.47], t(2135) = 6.53, p < .001) and outgroup members (Leave supporters: b = 9.36, 95% CI = [4.11, 14.60], t(1560) = 5.24, p < .001, Remain supporters: b = 17.03, 95% CI = [11.78, 22.28], t(1560) = 9.54, p < .001) perceived own-group members as feeling worse than other-group members in response to equivalent group-relevant experiences. By contrast, participants' predictions concerning the influence of targets' group category membership on ingroup and outgroup members' judgments of group-irrelevant emotions did not significantly deviate from how sampled ingroup and outgroup members actually responded (all ps > .06).

Comparing acontextualised and contextualised perceptions of empathic norms. To compare the two types of empathic norm perceptions (research question d), I regressed combined acontextualised and contextualised perceptions onto fixed effects for norm perception (levels: acontextualised, contextualised), empathiser (levels: average ingroup member, average outgroup member) and target (levels: own-group member, other-group member) separately for group-irrelevant and group-relevant experiences. In both models, I included random intercepts for participants and a correlated random slope for norm perception within participant.

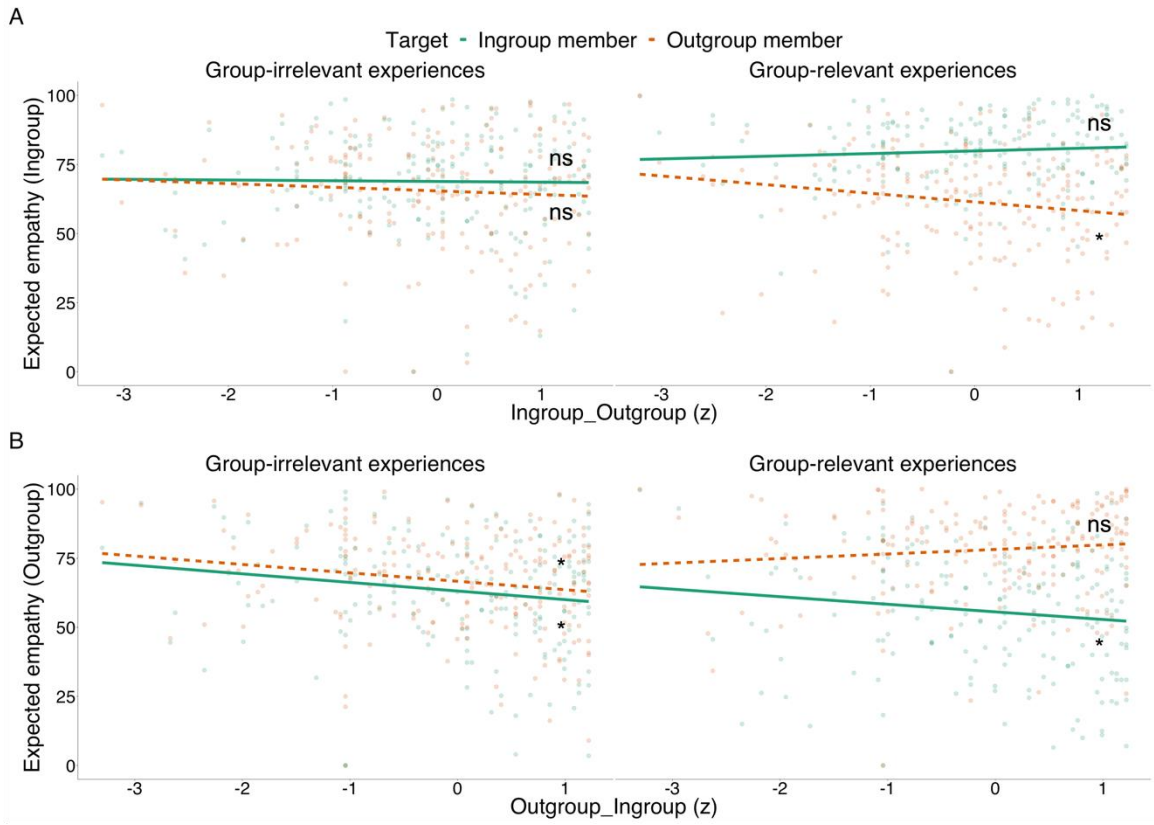
As can already be seen in Figure 3.4, contextual information and empathic targets' social identity salience exerted their influence in opposite directions. In comparison to acontextualised perceptions of empathic norms (expected intergroup empathy bias: $b = 23.41$, 95% CI = [21.01, 25.80], $z = 19.16$, $p < .001$), group membership's effect on perceptions of empathic norms was reduced when the provided contextual information did not make empathic targets' Brexit identities salient (expected intergroup empathy bias: $b = 3.65$, 95% CI = [2.45, 4.85], $z = 5.98$, $p < .001$). However, when contextual information highlighted empathic targets' Brexit identities, debiasing effects were lost (expected intergroup empathy bias: $b = 20.50$, 95% CI = [19.20, 21.90], $z = 17.05$, $p < .001$).

Relationship between perceptions of group relations and perceptions of emotion-perception and empathic norms. Finally, to examine whether general perceptions of intra- and intergroup group relations, embedded in prejudice perceptions, specifically predict perceptions of empathic norms, but not emotion-perception norms (research question f), as I observed in Experiment 4, I again adopted two approaches: (a) mixed effects modelling approach and (b) network modelling approach. Under the mixed effects modelling approach, I fitted linear mixed models separately for emotion-perception and empathic norms, and within each norm, separately for the average ingroup and the average outgroup member (empathiser). I always started from the baseline model which included target (levels: own-group member, other-group member) and experience type (levels: group-irrelevant, group-relevant) as fixed factors of interest. Building on the baseline model, I included prejudice perceptions in relevant directions and compared the more complex models with the preceding nested models using information criteria (AIC, BIC). In the main text, I report only results from the best performing models, while model selection is explained in more detail in Appendix 3.3.

As in Experiment 4, the best performing models included intergroup prejudice perceptions, but their prediction of perceived empathic norms was restricted to group-relevant experiences. Specifically, the more negative feelings participants thought the ingroup and the outgroup had towards each other in general, the less empathy (ingroup empathic norms: b = -3.13, SE = 1.28, t = -2.44, p = .019; outgroup empathic norms: b = -3.12, SE = 1.32, t = -2.35, p = .020) they expected for specific empathic targets' group-relevant experiences in the same direction (see Figure 3.5B). In the domain of group-irrelevant experiences, prejudice perceptions were either not predictive of empathic norm perceptions or did not discriminate between ingroup and outgroup targets (see Figure 3.5A). Finally, including prejudice perceptions as predictors either did not improve model fit in the domain of emotion perception or did not interact with other factors of interest (see Appendix 3.3).

Figure 3.5

Perceptions of empathic norms in relation to intergroup prejudice perceptions (Experiment 5)



Notes. The prejudice perception scales were reversed and standardised such that higher values indicate that (A) participants believed ingroup members harboured more negative feelings towards the outgroup in general, and (B) outgroup members harboured more negative feelings towards the ingroup in general (r = .48, p < .001). In Experiment 5, since the correlation between two reciprocal prejudice perceptions included in A and B was significantly lower than in Experiment 4, I did not statistically control for the irrelevant, but correlated perception to produce the visualisations and statistical output reported in the text, but the results did not change even after controlling for it. ns = Slope not significant. * Slope significant at p < .05.

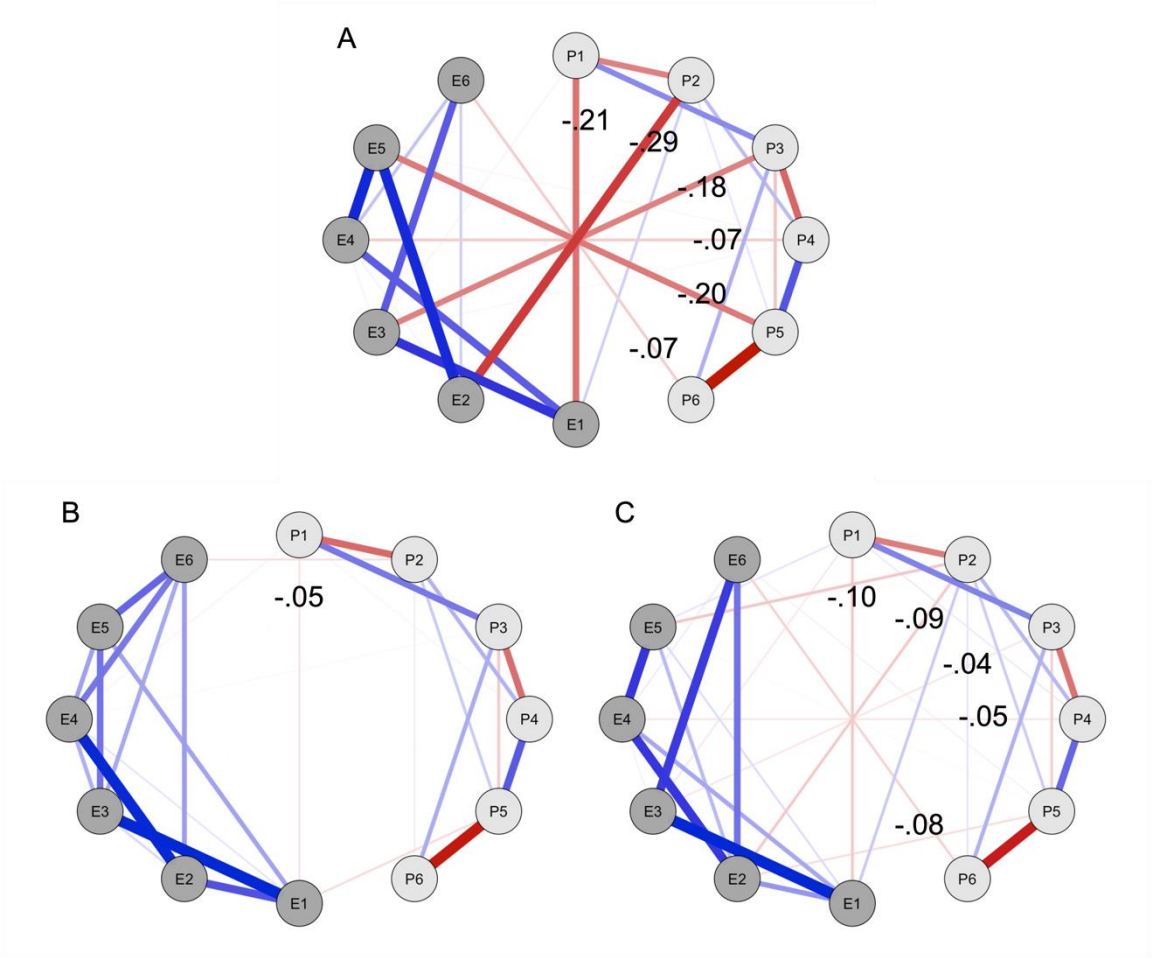
Under the network modelling approach, as in Experiment 4, I fitted undirected partial correlation networks which always included own prejudices, and prejudice perceptions on one side. On the other side, separate networks included one of the following: (a) acontextualised perceptions of trait empathy and empathic norms (Figure 3.6A), (b) contextualised empathy and perceptions of empathic norms averaged across empathic targets'

group-irrelevant experiences (Figure 3.6B), and (c) contextualised empathy and perceptions of empathic norms averaged across empathic targets' group-relevant experiences (Figure 3.6C). Accuracy of estimated network connections is reported in Appendix 3.4.

In terms of results, the same pattern of cross-domain connections as in Experiment 4 was identified when acontextualised perceptions of trait empathy and empathic norms (see Figure 3.6A) and contextualised empathy and perceptions of empathic norms for targets' group-relevant experiences were included as nodes (see Figure 3.6C). No reliable cross-domain connections were observed in the network which included contextualised empathy and perceptions of empathic norms for targets' group-irrelevant experiences (see Figure 3.6B). Following my initial causal reasoning, this might suggest that participants primarily used their perceptions of general intra- and intergroup relations to make empathic predictions in contexts where these perceptions were most diagnostic and relevant (i.e., in contexts where Brexit identities were salient).

Figure 3.6

Empathy and perceptions of empathic norms in relation to prejudices and prejudice perceptions (Experiment 5)



Notes. Regularised partial correlation networks based on (A) participants' acontextualised perceptions of ingroup and outgroup members' trait empathy and empathic norms, (B) participants' contextualised empathy and perceptions of empathic norms for empathic targets' group-irrelevant experiences, and (C) participants' contextualised empathy and perceptions of empathic norms for empathic targets' group-relevant experiences. E1 = Perceptions of ingroup members as empathic (A) or one's own empathy for ingroup targets (B-C). E2 = Perceptions of outgroup members as empathic (A) or one's own empathy for outgroup targets (B-C). E3 = Expected empathy of ingroup members for ingroup targets. E4 = Expected empathy of ingroup members for outgroup targets. E5 = Expected empathy of outgroup members for ingroup targets. E6 = Expected empathy of outgroup members for outgroup targets. P1 = One's own prejudice towards ingroup targets. P2 = One's own prejudice towards outgroup targets. P3 = Expected prejudice of ingroup members towards ingroup targets. P4 = Expected prejudice of ingroup members towards outgroup targets. P5 = Expected prejudice of outgroup members towards ingroup targets. P6 = Expected prejudice of outgroup members towards outgroup targets. Displayed values represent strengths of significant edges.

Discussion

Experiment 5 was designed to examine whether the findings from Experiment 4, which focused on the post-conflict context of Croatia and Serbia, replicate in another competitive intergroup context – the political intergroup context of Brexit – and to assess whether the group-relevance of emotional experiences affects one's own emotion perception and empathy as well as perceptions of respective emotion norms. First, Experiment 5 replicated Experiment 4's finding that people tend to see other ingroup and outgroup members as more biased in their emotion perception and empathy than they are themselves (research question b). This strengthens my conclusion that these self-enhancing perceptions of emotion-perception and empathic norms are not context-specific and apply across different competitive intergroup environments.

Second, expanding on Experiment 4's findings, I identified circumstances in which intergroup empathy biases and overly negative perceptions of emotion-perception and empathic norms are particularly likely to occur, specifically when empathic targets' emotional experiences make their social identities more salient. These findings offer a more nuanced view of contextual information's ability to decrease intergroup bias in empathy and in perception of emotion-perception and empathic norms, suggesting that providing contextual information about specific empathic targets and their emotional circumstances might only have debiasing effects when the contextual information does not highlight divisive social identities.

However, increased salience of social identity might not be the only relevant factor in explaining these effects. It is also possible that participants' responses were affected by (a) how much participants themselves thought empathic targets' attribution of personal experiences to Brexit-related concerns was justified and (b) how much participants assumed others (average ingroup and outgroup members) thought empathic targets' attribution of

personal experiences to Brexit-related concerns was justified. To illustrate, participants could have shown more empathy for ingroup than outgroup Brexit-related emotional experiences because they perceived political ingroup members' attribution of these emotional experiences to Brexit-related concerns as more justified than the equivalent attribution of political outgroup members. Similarly, participants could have assumed that average ingroup and outgroup members would show more empathy for own-group than other-group empathic targets because participants thought that average ingroup and outgroup members would perceive own-group targets' attribution of personal experiences to Brexit as more justified than other-group targets' attribution. Future research should therefore assess the extent to which: (a) participants themselves perceived the empathic targets' attribution of their emotional experiences to Brexit as justified; (b) participants expected others to perceive the empathic targets' attribution as justified; and (c) these judgments of attribution justifiability influenced people's own empathy and perceptions of emotion-perception and empathic norms. In summary, it is important to delineate the relative influence of targets' social identity salience and judgments of targets' attribution of emotional experiences on emotion perception, empathy, and perceptions of respective norms.

Third, I again found evidence of empathic norm perception asymmetry but only when empathic targets' social identities were particularly salient (i.e., when participants reported their acontextualised perceptions of empathic norms or their contextualised perceptions of empathic norms in the domain of group-relevant emotional experiences). In contrast to Experiment 4, where norm perception asymmetry was restricted to perceptions of empathic norms, Experiment 5 suggested that people might assume that outgroup members would be more affected by empathic targets' group category membership than ingroup members not only in the way they respond to emotions, but also in the way they perceive them. However, the fact that norm perception asymmetry was more robust in the domain of empathy, which is

commonly perceived as morally and socially valuable (Fowler et al., 2021), than in general emotion perception might be seen as indicating that the more morally right a judgment or an emotion is perceived to be, the more likely the people are to display own-group-favouring asymmetry. Nonetheless, this remains to be empirically tested and future research should examine the boundary conditions of norm perception asymmetry, clarifying whether outgroup members are expected to be more biased than ingroup members across a wider range of intergroup judgments, emotions, and behaviours with connotations of morality.

Finally, general perceptions of intra- and intergroup relations' valence (*prejudice perceptions*) closely tracked same-direction perceptions of empathic, but not emotion-perception norms, in contexts where it was particularly relevant to think of empathy or a lack thereof as a product of group relations. This stable pattern of results highlights the importance of clarifying the relationship between (perceptions of) group relations and (perceptions of) empathy, both in terms of how (perceptions of) group relations can influence (perceptions of) empathy and in terms of how witnessing intergroup empathy or a lack thereof can inform inferences about underlying group relations.

Experiment 6a

Although Experiments 4 and 5 examined vastly different intergroup contexts, these contexts still share some features that have been at the core of social psychological theorising; notably, (past or present) intergroup competition and conflict. Experiment 6a, thus, sought to assess to what extent the consistent effects documented in Experiments 4 and 5 reflect participants' abstract intuitions about how group category membership itself affects empathy (e.g., "people always feel more empathy for others who belong to their group") rather than their more specific perceptions about how the particular sampled groups operate and relate to one another. To answer this question, Experiment 6a investigated whether creating minimal groups is sufficient to produce people's tendency to assume the existence of

intergroup empathy bias in others (research question a) as well as their tendency to assume that outgroup members are more subject to intergroup empathy bias than ingroup members (*empathic norm perception asymmetry*, research question d)[61].

Importantly, nowhere in Experiment 6a was the reason for minimal group assignment given to participants. However, as suggested in Chapter 2, this does not prevent participants from automatically assuming certain, usually coalitional, relations within and between minimal groups (Balliet et al., 2014; Brewer, 2008; Yamagishi & Kiyonari, 2000). Therefore, even though I use the standard term "minimal groups" in my description and discussion of results, I do not adhere to the view that these groups reflect blank slate group categories devoid of any relational meaning. Nonetheless, in contrast to groups examined in Experiments 4 and 5, minimal groups are certainly free from years of intergroup competition and conflict and thus provide a way to assess people's assumptions about how groups affect others' empathy in the absence of features relating to their specific history and characteristics.

Method

Participants. Eligible participants had to reside in the UK, have no formal education in psychology, and have no history of diagnosed mental health disorders or autism. A total of 352 participants were recruited through social media and university recruitment pools and completed a two-part experiment in Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). In this chapter, I will only present data from participants allocated to the control condition in

[61] It is important to note that research questions a and d were only a subset of research questions that Experiment 6 aimed to address but are the only relevant research questions in the context of this chapter. Thus, in this chapter, I will primarily present aspects of the research design that are of relevance to these two research questions under Experiment 6a while other research questions will be covered in the next chapter, under Experiment 6b.

the second part of the experiment (N = 118; 47 men, 69 women, 2 non-binary; M_{AGE} = 30.77, SD_{AGE} = 10.28) since this was the only condition that allowed me to estimate participants' subjective perceptions of empathic norms unaltered by experimental input. No participants were excluded. Experiment 6a had 80% power to detect a small effect size equivalent to d = 0.24 (Faul et al., 2009). Post hoc sensitivity simulations (1000 iterations) for the fitted linear mixed model (Green & MacLeod, 2016) showed that our observed estimate was 62.6% larger than the minimal detectable estimate with 80% power.

Procedure. In the first part of the experiment, participants were randomly allocated to a minimal group (Square or Triangle) and asked to describe one negative personal experience. Participants were told that in the week between the two parts of the experiment, collected personal experiences would be shown to randomly selected participants from both minimal groups, to assess how people perceive experiences of strangers and how much empathy they tend to show for them. In reality, no participants were exposed to any personal experiences during the week between the two parts of the experiment. This cover story was used only to explain the availability of supposed feedback information about outgroup intergroup empathic norms in the second part of the experiment (relevant for Experiment 6b in Chapter 4).

After a week, the same participants completed the second part of the experiment, wherein they were randomly assigned to one of the following conditions: (a) impartial outgroup empathy condition, (b) biased outgroup empathy condition, and (c) control condition. The first two conditions (a and b) included explicit feedback about minimal outgroup members' supposed empathy for participants' minimal ingroup and therefore do not help to assess people's unmanipulated perceptions of empathic norms. I therefore focus on the control condition (c), which made no attempt to manipulate participants' perceptions of empathic norms. Instead, participants immediately proceeded to the task where they were

presented with pre-selected short stories – eight describing negative events supposedly reported by members of their minimal ingroup, and eight describing negative events supposedly reported by members of their minimal outgroup. In reality, these 16 short stories were not produced by real participants but were instead chosen based on a separate pilot study. More details about the pilot study can be found in the Appendix 3.2, and all stories can be found on Open Science Framework under Experiment 3 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d). Finally, unlike in Experiments 4 and 5, participants in Experiment 6a provided emotion perception and empathy ratings only from their own perspective (emotion perception: *How bad/good do you think this person felt?;* empathy/schadenfreude: *How bad/good does what happened to this person make you feel?*).

After completing the task, participants in the control condition were asked to estimate how much an average minimal ingroup member, and an average minimal outgroup member, who were allegedly exposed to the same personal experiences in the week between the two parts of the experiment, empathised with ingroup and outgroup targets. Participants gave their responses on 0-100 slider scales, where 0 meant "The average [minimal ingroup/outgroup member] felt no empathy" and 100 meant "The average [minimal ingroup/outgroup member] felt total empathy". These responses capture participants' acontextualised perceptions of ingroup and outgroup empathic norms in a minimal intergroup context.

Results

I regressed acontextualised perceptions of empathic norms onto fixed effects for empathiser (levels: average ingroup member, average outgroup member) and target (levels: own-group target, other-group target) and included a random intercept for participants and a correlated random slope for empathiser within participant. Separately, I regressed trial-by-

trial empathy responses in the task onto fixed effect for target (levels: own-group target, other-group target) and included a random intercept for participants.

I observed a main effect of targets' group category membership ($F(1, 233.99) = 48.86$, $p < .001$). This meant that, similar to the real-world contexts, participants expected both an average ingroup and an average outgroup member to empathise more with own-group targets than with other-group targets ($b = 6.89$, SE = 0.96, 95% CI = [4.94, 8.83], $t(234) = 6.99$, $p < .001$). Unlike in the real-world intergroup contexts sampled in Experiments 4 and 5, however, there was no significant two-way interaction between empathic targets' group membership and empathiser ($F(1, 233.99) = 0.05$, $p = .827$). Finally, unlike acontextualised perceptions of empathic norms, participants' trial-by-trial empathy responses during the task were not significantly influenced by empathic targets' group category membership ($b = -0.07$, SE = 1.13, 95% CI = [-2.29, 2.16], $t(1769) = -0.06$, $p = .953$).

Discussion

Experiment 6a used the minimal groups paradigm to eliminate the influence of prominent confounding factors in Experiments 4 and 5 (e.g., history of competition and conflict) and thus, better isolate the effects of people's assumptions about how sharing group category membership influences others' empathy in its own right (Cikara & Van Bavel, 2014; Dunham, 2018). The findings provided support for the idea that minimal groups (with their meaningful relational implications) are sufficient to produce acontextualised expectations of intergroup empathy bias but are less likely to produce the empathic norm perception asymmetry identified in Experiments 4 and 5. It is important to note, however, that the acontextualised intergroup empathy bias participants expected in the minimal intergroup environment was less pronounced than the acontextualised intergroup empathy bias participants expected in the real-world intergroup environments examined in Experiments 4 and 5 (see Appendix 3.5). This suggests that features associated with the real-

world intergroup contexts sampled in Experiments 4 and 5 (e.g., history of conflict, competition) may have amplified biased perception of empathic norms and led to its asymmetry.

General discussion

In this chapter, I have presented the findings from three experiments investigating people's subjective perceptions of empathic norms in intergroup contexts, the accuracy of these perceptions and their dependence on contextual information about empathic targets and their emotional experiences as well as on individual differences in how relevant group relations are perceived. In accordance with initial hypotheses, all three experiments provided convergent evidence that perceptions of empathic norms are overly negative and inaccurate in intergroup contexts, with participants seeing ingroup and outgroup members as more biased in their allocation of empathy than participants are themselves and than sampled ingroup and outgroup members turned out to be. Participants also tended to assume that both ingroup and outgroup members would perceive own-group empathic targets as experiencing more negative emotions than other-group empathic targets in response to equivalent events. Comparatively speaking, assumed intergroup bias in empathy was consistently larger than assumed intergroup bias in emotion perception.

In addition, participants judged that competitive real-world – ethnic (Experiment 4) and political (Experiment 5) – but not minimal (Experiment 6a) outgroup members would discriminate more in their intergroup empathy than ingroup members. This *empathic norm perception asymmetry* was mainly observed when empathic targets' social identities were salient, and specifically (a) when perceptions of empathic norms were reported in the absence of contextual information so that empathisers and empathic targets were primarily viewed in terms of their specific social identities (e.g., Experiment 4: "Croatian person", "Serbian person"; Experiment 5: "Leave supporter", "Remain supporter"), or (b) when perceptions of

empathic norms were reported in the presence of contextual information that made empathic targets' social identities salient. However, the fact that I observed empathic norm perception asymmetry even for emotional experiences that were unrelated to empathic targets' ethnic identities and experience of conflict (Experiment 4) suggests that the asymmetry is not restricted to contexts with high social identity salience. However, it remains possible that the experimentally manipulated cues (e.g., country name, flag, and ethnic name) used to inform participants of the empathic targets' group category membership were themselves sufficient to increase social identity salience, even when empathic targets' emotional experiences were group-irrelevant. Finally, it is worth noting that Experiment 5 unexpectedly suggested that participants might also assume that empathic targets' group membership has a greater influence on perceptions of their emotions for outgroup members than it does for ingroup members. An important factor that may make empathy particularly prone to norm perception asymmetry in comparison to general emotion perception is its assumed link with morality (Fowler et al., 2021; Schumann et al., 2014). Future research should seek to clarify the role of morality and outline the boundary conditions of norm perception asymmetry by investigating whether similar asymmetry arises for other morally loaded intergroup judgments, emotions, and behaviours. Additionally, I invite future researchers to examine perceptions of empathic norms in different intergroup contexts, especially in contexts which are less characterised by competition, moralisation, and conflict than those focused on in the present work (see Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). The current chapter has taken the first step towards diversifying examined intergroup contexts by contrasting minimal group findings with findings from two real-world competitive intergroup contexts, but systematically expanding this research so that it samples more cooperative intergroup contexts and manipulates intergroup relations experimentally will help to increase the confidence that

features such as intergroup competition are necessary for the empathic norm perception asymmetry observed in the current experiments.

Furthermore, I speculate that ingroup favouritism, wherein the ingroup is seen more favourably than the outgroup (including in terms of morality – Enock et al., 2021; Liberman et al., 2018), might be generating empathic norm perception asymmetry through own-group love/other-group hate motive attribution asymmetry (Waytz et al., 2014). Specifically, direct comparison of expected empathic responses to group-irrelevant and group-relevant emotional experiences in Experiment 5 raises the possibility that the ingroup's assumed intergroup empathy bias is seen as arising primarily from own-group love, but not other-group hate, whereas the outgroup's assumed intergroup empathy bias is more likely to be seen as a product of both own-group love and other-group hate. In other words, outgroup members' intergroup empathy bias might be seen as driven by two motives instead of one, which could explain why it is assumed to be greater than ingroup members' empathy bias. In both Experiments 4 and 5, participants' responses reflected expectations that outgroup members would show lower empathy for other-group empathic targets but not higher empathy for own-group empathic targets in comparison to ingroup members. This pattern of results is consistent with the idea that an own-group love/other-group hate motive asymmetry underlies the observed empathic norm perception asymmetry. Further in line with this interpretation, perceptions of ingroup and outgroup empathic norms were not asymmetrical in a minimal intergroup context where own-group-favouring biases are thought to reflect own-group love, but not other-group hate (Molenberghs et al., 2014; Molenberghs & Louis, 2018). However, future studies should directly assess own-group love and other-group hate motive attribution in order to substantiate this interpretation.

More work is also needed to address the consequences of negatively biased perceptions of empathic norms, examining how perceived empathic norms inform personal

empathy for outgroup members, and how they impact reactions to actual displays of intergroup empathy and prosocial behaviour (Borinca et al., 2021b). Additionally, while the current set of experiments only addressed subjective perceptions of descriptive empathic norms – beliefs about how much empathy relevant others feel for own-group versus other-group targets – future research should also assess subjective perceptions of injunctive empathic norms – beliefs about how much relevant others approve of empathy for own-group and especially other-group targets (Cialdini, 2003, see Fowler et al., 2021). This line of work is particularly important since normative influence is most likely to encourage or discourage empathy and prosociality when (perceptions of) descriptive and injunctive empathic norms are aligned (Weisz et al., 2021). Finally, since previous research has focused on the influence of ingroup and outgroup intergroup empathic norms in isolation (e.g., Gubler et al., 2015; Tarrant et al., 2009), it remains an open question how people would respond when both kinds of empathic norms are available simultaneously and are either aligned or misaligned (see Romano & Balliet, 2017). The next chapter in this thesis provides the first known empirical evidence pertaining to this research question. Overall, answering all these questions should lead to a better understanding of intergroup empathy with respect to the role of empathic norms.

Furthermore, perceptions of empathic norms also varied depending on the availability of contextual information about empathic targets' emotional experiences and the group relevance of these emotional experiences. Providing context about specific empathic targets and their emotional experiences substantially reduced but did not fully remove bias in perceptions of empathic norms as intergroup biases continued to be expected from others even when context, provided in the form of group-irrelevant emotional experiences, facilitated construal of empathic targets beyond their social identities At the same time, these results suggest that offering contextual information describing group members' idiosyncratic

emotional circumstances only has debiasing effects when this context does not enhance divisive social identities. Specifically, when contextualised experiences were unrelated to the empathic targets' social identities (i.e., in group-irrelevant emotional experiences), participants showed no significant intergroup empathy bias and expected only a slight bias from average ingroup and outgroup members (i.e., across experiments, the average bias in group-irrelevant, contextualised perceptions of empathic norms was equal to a 3- to 7-point difference on a 100-point scale in contrast to a 17- to 28-point average bias in acontextualised perceptions of empathic norms). On the other hand, when contextualised experiences increased the salience of empathic targets' social identities (i.e., in group-relevant emotional experiences), actual intergroup empathy bias was displayed (i.e., the average actual bias was equal to a 7-point difference on a 100-point scale) and almost unanimously expected from average ingroup and outgroup members (i.e., the average bias in group-relevant, contextualised perceptions of empathic norms was equal to a 18- to 22-point difference on a 100-point scale). Of note, I acknowledge that my group-relevant stimuli might have not only made empathic targets' social identities salient but also created room for participants to assess the justifiability of empathic targets' responses to their group-relevant experiences (e.g., participants could have perceived an empathic target's decision to vote Remain as a justified or an unjustified response to losing an opportunity to participate in the EU exchange programme). Thus, it is possible that greater actual and expected intergroup empathy biases in the group-relevant condition were at least partially driven by underlying justifiability judgments in addition to empathic targets' social identity salience. However, it remains for future research to delineate the role of these two factors.

The finding that group-relevant experiences led to greater actual and expected intergroup empathy bias may seem inconsistent with recent evidence that political opponents' opinions garner greater respect when based on relevant personal experiences rather than facts

(Kubin et al., 2021). First, it is plausible that relevant personal experiences evoke respect, operationalised as a more distanced response of consideration and tolerance, in the absence of empathy (e.g., *I respect where they are coming from, but I do not resonate with them.*). Second, previous work (Kubin et al., 2021) has focused on experiences surrounding opinions which could be considered somewhat diagnostic of dominant political identities in the US (opinions on abortion or taxation) but are unlikely to constitute salient social identities in themselves[62]. In contrast, Experiment 5 focused on Brexit labels (Leave and Remain), which are issue-based opinions at their core, but are also more likely to be viewed as social identities, similar to "Democrats" and "Republicans" in the US (Hobolt et al., 2021). Since social identities have a plethora of consequences that issue-based opinions might lack, it is important to make a clearer distinction between identities and opinions in future research, but also to engage in theorising about the consequences of issue-based opinions becoming self-standing social identities (e.g., "believing that the UK should remain in the European Union" versus "identifying as a Remainer"). For these reasons, I encourage researchers to investigate the effects of sharing both identity- and opinion-relevant experiences on a wider variety of outcomes.

Finally, my results emphasised the relationship between prejudice perceptions – beliefs about how positively or negatively group members feel towards their own groups and each other – and perceptions of empathic norms. For instance, the more people believed that ingroup and outgroup members disliked each other, the less empathy they expected in the same direction. This was again true primarily when empathic targets' social identities were made salient by either their emotional experiences or experimentally manipulated cues,

[62] It is reasonable to assume that opinions concerning abortion evolved into moralised social identities in the wake of the Supreme Court ruling on Roe v. Wade in the United States. Hence, it would be important to examine whether sharing personal experiences about this political issue is still likely to help foster respect even after Supreme Court's decision.

making it particularly relevant to think of empathy or a lack thereof as a product of group relations. While in the context of Experiments 4 and 5, I assumed that participants relied on their perceptions of known groups' relations to generate empathic predictions in previously unseen social situations, more broadly, I theorise that this relationship is likely bidirectional. Specifically, people might first observe empathic displays of others and infer valence of intra- and intergroup relations (see Thornton et al., 2019), captured in perceptions of ingroup and outgroup prejudices, which can then act as a guiding framework when predicting group feelings and behaviours in relevant circumstances. Note that valence here is a dimension of group relations rather than traits, reflecting perceptions of how groups relate to each other, and not what they are separately like. Thus, it is important to consider expanding current models of intergroup, and likely also interpersonal, social perception to capture perceptions not only of own and other groups' traits but also of relations within and between different groups (and individuals).

In summary, the experiments described in this chapter extend existing research on intergroup empathy by showing that people hold overly negative perceptions of empathic norms in intergroup contexts, consistently expecting preferential intra- in comparison to intergroup empathy from both ingroup and outgroup members. Across different assessments of empathic norm perceptions and across different types of empathic targets' emotional experiences, the average intergroup empathy bias that was expected from others ranged from a 3-point to a 28-point difference on a 100-point scale. Thus, an important implication of current experiments is that while perceptions of empathic norms in intergroup contexts tend to be characterised by assumptions of intergroup empathy bias, the size of this assumed intergroup empathy bias can vary dramatically depending on the nature of empathy-evoking emotional content and the perceptions of relevant group relations. Furthermore, not only do perceptions of empathic norms differ from the actual degree of intergroup empathy bias

observed, but they are also asymmetrical in competitive real-world intergroup contexts, with outgroup members assumed to discriminate more in their empathy than ingroup members. Finally, perceptions of ingroup and outgroup prejudices reliably predict same-direction perceptions of empathic norms, highlighting the need to expand the focus to cover people's perceptions of within- and between-group relations in addition to their perceptions of groups' traits.

Despite intergroup empathy not being in such short supply as is subjectively perceived, overly negative perceptions of empathic norms might themselves represent a barrier to conflict resolution. For example, they might limit individuals' willingness to attend to outgroup members' emotions, as little reciprocity is expected, or taint their interpretations of actual displays of intergroup empathy, leading to attributions of insincerity and constraining any positive influence of these displays on outgroup-oriented feelings and behaviours. Acknowledging the importance of perceptions of normative standards concerning feelings and behaviours in intergroup contexts, especially when these perceptions might not accurately reflect reality, will help to inform our understanding of the psychological underpinnings of long-standing intergroup conflict and will contribute to the principled improvement of conflict-reduction interventions.

Building on the growing literature emphasising the primacy of norm *perceptions* over actual norms and even personal beliefs in shaping behaviour (e.g., Jachimowicz et al., 2018; Tankard & Paluck, 2016; Vallentin-Holbech et al., 2017), Chapter 3 sought to understand people's unaltered *perceptions* of empathic norms in intergroup contexts. Nevertheless, the experiments reported in Chapter 3 did not experimentally manipulate intergroup empathic norms, thereby not allowing firm conclusions about how these norms affect one's own empathy for outgroup members' emotions. In the next chapter, I revert to this causal question, which has been at the centre of most published work on empathic norms, to better

understand the pathways through which intergroup empathic norms encourage empathy across group lines as well as how people navigate social environments where multiple – congruent, or incongruent – intergroup empathic norms are simultaneously active.

Open practices

The pre-registration for Experiment 4 can be found at https://osf.io/e96k3/?view_only=f5240af38e964a988f71b5111b799d7d, while the pre-registration for Experiment 5 can be found at https://osf.io/kw2m6/?view_only=3c0fbb21c3604198aec3c1b83345b479. Materials, data, and analysis code for all three experiments are available at https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d. Finally, research described in Chapter 3 is currently available as a pre-print and can be found at https://psyarxiv.com/xj9bv/.

CHAPTER 4: LEARNING TO EMPATHISE: HOW DO INTERGROUP EMPATHIC NORMS INFLUENCE OUTGROUP-ORIENTED EMPATHY?

Abstract

Although people often feel reduced empathy for outgroups, intergroup empathic norms can help increase it. For instance, people empathise more with outgroup members when their ingroup members are also empathetic (conformity to ingroup intergroup empathic norms) or when outgroup members are empathetic towards one's ingroup (reciprocity of outgroup intergroup empathic norms). However, little is currently known about the mechanisms through which intergroup empathic norms achieve their influence or about preferences people display when faced with ingroup and outgroup intergroup empathic norms that demand misaligned emotional reactions (e.g., ingroup norms demand apathy while outgroup norms demand empathy). In response to the first question, in Experiment 6b (N = 333), I examined whether outgroup intergroup empathic norms encourage empathy across group lines by reducing how much outgroup members are subsequently perceived as deserving of their adversities. I found no evidence that outgroup intergroup empathic norms change appraisals of outgroup members' deservingness, but due to several methodological limitations, these appraisals should not be discarded as a candidate mechanism. In response to the second question, Experiment 7 (N = 251) investigated whether people preferentially conform to ingroup intergroup empathic norms or reciprocate outgroup intergroup empathic norms. Specifically, Experiment 7 focused on displays of intergroup empathy and tracked how they vary across contexts where both ingroup and outgroup intergroup empathic norms are available simultaneously and where conformity and reciprocity promote congruent or incongruent emotional reactions (i.e., apathy versus empathy). I found that reciprocity

enhanced displays of empathy for outgroup members more than conformity. However, participants did not form perceptions of intergroup empathic norms in an unbiased manner, but most frequently erred towards judging that ingroup and outgroup members displayed less intergroup empathy than they in fact did. Together, these findings suggest that outgroup intergroup empathic norms might be a more important driver of people's motivation to display empathy for outgroup members than ingroup intergroup empathic norms, but more work is needed to assess the generalisability of this result and explain the exact mechanisms through which intergroup empathic norms shape displays and feelings of intergroup empathy.

Introduction

In Chapters 1 and 3, I reviewed existing evidence documenting that intergroup empathy can be encouraged, at least in the short term, by portraying it as normative among relevant others. However, I then emphasised that outside of psychological interventions and experiments, operative intergroup empathic norms are rarely signalled in such an explicit manner. Instead, people more often form perceptions of intergroup empathic norms through more subtle experiences and observations over time and subsequently rely on these perceptions to navigate intergroup interactions requiring empathy. Expanding on the growing literature emphasising the primary role of norm *perceptions* in shaping behaviour, Chapter 3 correspondingly focused on people's unaltered *perceptions* of empathic norms in intergroup contexts. Across three experiments, spanning ethnic, political, and minimal groups, perceptions of intergroup empathic norms were found to be overly negative and inaccurate but simultaneously context-dependent, varying as a function of the characteristics of empathic targets and emotional experiences as well as perceivers' representations of group relations.

Nevertheless, given that experiments described in Chapter 3 did not systematically manipulate intergroup empathic norms, they could not contribute to answering the question of how these norms influence one's own empathy for outgroup emotions. This causal question, which has propelled most scientific discussion on intergroup empathic norms, is therefore addressed in the current chapter in the context of two experiments – Experiment 6b and Experiment 7. First, in Experiment 6b, I examined whether outgroup intergroup empathic norms encourage empathy across group lines by reducing how much outgroup members are subsequently perceived as responsible for and deserving of their adversities. Second, recognising that past work has only looked at the influence of ingroup and outgroup intergroup empathic norms one at the time, in Experiment 7, I investigated how much

intergroup empathy people display across social environments where cues of both ingroup and outgroup empathic norms are available simultaneously (i.e., both ingroup and outgroup members are present) and either aligned or misaligned (i.e., both promote empathy or only one promotes empathy and the other one promotes apathy). In other words, Experiment 7 sought to ascertain whether people preferentially *conform* to ingroup intergroup empathic norms or *reciprocate* outgroup intergroup empathic norms. Finally, as in Chapter 3, the term *empathy* refers to the vicarious adoption of empathic targets' emotions (i.e., "feeling with" the empathic targets), while the term *empathic norms* refers to social norms which signal how much relevant others – ingroup and outgroup members – vicariously adopt empathic targets' emotions.

Akin to previous empirical chapters, this chapter is divided into two sections, each corresponding to a single experiment. Within each section, I provide experiment-specific rationale, hypotheses, methods, results, and discussion. I conclude with a general discussion, in which I do not focus on the details of each experiment, but rather describe their combined implications and highlight avenues for future research.

Experiment 6b

What determines whether we respond with empathy, apathy, or schadenfreude to social group members' misfortunes? For instance, why are people likely to respond with empathy after seeing communities struggling with floods (e.g., Zaki, 2020), and with apathy or even schadenfreude after seeing hedge funds shutting down and their employees losing their jobs in the GameStop saga (e.g., Menton, 2021)? In response to this question, previous research has proposed that whether we perceive the specific misfortune as deserved or undeserved (Feather, 2006) is one of the key factors shaping our emotional responses. Specifically, if group members' misfortune is perceived as deserved, people are more likely to feel schadenfreude in response (Feather et al., 2013; Smith & van Dijk, 2018; Smith et al.,

2009; see also Li et al., 2019), and if it is perceived as undeserved, empathy and compassion are more common emotional responses (Feather, 2006; Hudson, 2022; see also Berndsen et al., 2017). Thus, we could say that people are likely to feel empathy for communities battling floods because they perceive these communities as blameless and undeserving of their misfortunes. In contrast, people are likely to feel apathy or schadenfreude for hedge fund employees because they perceive the employees' misfortunes as a result of greed, rendering the misfortunes deserved (Watanabe et al., 2022; but see Lucas & Kteily, 2018).

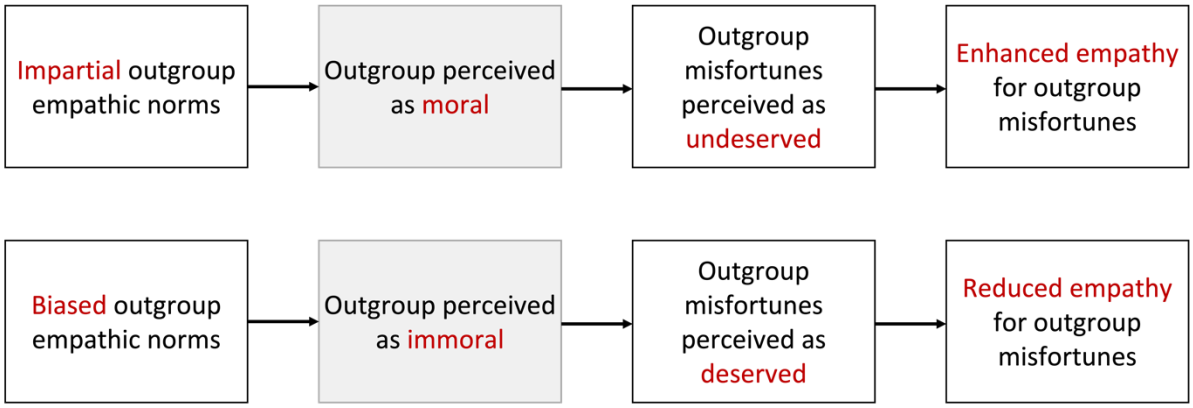
While there is widespread agreement that perceived deservingness is a crucial determinant of whether someone's misfortunes generate empathy or schadenfreude, no research to date has examined its role in the context of different intergroup empathic norms. Bridging two lines of research, Experiment 6b sought to examine whether outgroup intergroup empathic norms exert their influence on personal empathy for outgroup members, as was documented in previous literature (Gubler et al., 2015; Hein et al., 2016; Nadler & Liviatan, 2006) because they alter subsequent cognitive appraisals of deservingness. I focused on outgroup intergroup empathic norms, operationalised through portraying outgroup members as either impartial or biased in their empathy, as these norms are particularly likely to shape deservingness judgments and corresponding emotional responses, due to being indicative of outgroup morality (Fowler et al., 2021). Specifically, finding out that outgroup members do or do not discriminate in their empathy based on group category membership should drive participants' inferences of lower or higher outgroup morality respectively, which should in turn influence how participants subsequently appraise and respond emotionally to outgroup members' misfortunes (see Figure 4.1). Given that immoral individuals' misfortunes are more likely to be perceived as deserved and to trigger schadenfreude (Berndsen & Tiggemann, 2020; Brambilla & Riva, 2017), I hypothesised that participants exposed to *biased* outgroup intergroup empathic norms, wherein outgroup

members display greater empathy for own-group than other-group targets, would subsequently construe outgroup misfortunes as more deserved and show reduced empathy for outgroup members in comparison to participants exposed to *impartial* outgroup intergroup empathic norms, wherein outgroup members display equal empathy for own-group and other-group empathic targets.

To disentangle whether the predicted difference between participants exposed to impartial and biased outgroup intergroup empathic norms stems from (a) impartial norms reducing attributions of deservingness and enhancing outgroup-oriented empathy, (b) biased norms enhancing attributions of deservingness and reducing outgroup-oriented empathy or (c) both, Experiment 6b also included a control condition, where no information about outgroup empathic norms was provided (Cikara et al., 2014). In addition to acting as a baseline, the control condition allowed the same hypotheses to be addressed from a different angle. Specifically, if the relationship between perceived outgroup intergroup empathic norms, perceived deservingness of outgroup misfortunes and outgroup-oriented empathy is adequately captured by the hypotheses listed above, participants' perceptions of outgroup intergroup empathic norms in the control condition should similarly predict how much they perceive outgroup misfortunes as deserved and how much empathy they report in response. To illustrate, participants who, in the absence of any explicit information, assume that outgroup members discriminate in their empathy in favour of their own group members should judge outgroup misfortunes as more deserved and show lower empathy for outgroup members in comparison to participants with more impartial perceptions of outgroup empathic norms. Finally, in Experiment 6b, these hypotheses were tested in the context of minimal groups (Otten, 2016; Tajfel et al., 1971), leaving it for future research to test them in real-world intergroup environments.

Figure 4.1

Proposed mediation model from (perceived) outgroup intergroup empathic norms to outgroup-oriented empathy



Notes. Attributions of outgroup morality were not explicitly assessed in Experiment 6b, indicated by the grey square in the figure. For simplicity, I illustrate the proposed mediation model in categorical terms (e.g., "moral" or "immoral", "deserved" or "undeserved"), but in reality, these judgments more likely vary in degree (e.g., from "not moral at all" to "extremely moral", from "not deserved at all" to "extremely deserved").

Finally, it is important to note that Experiment 6a, discussed in Chapter 3, and Experiment 6b were two parts of a single experiment which addressed complementary but distinct research questions. Specifically, Experiment 6a focused narrowly on the control condition to describe people's unaltered perceptions of intergroup empathic norms in the context of minimal groups. In contrast, Experiment 6b presents all three experimental conditions – (a) impartial outgroup empathy condition, (b) biased outgroup empathy condition and (c) control condition – in more detail and tests whether outgroup intergroup empathic norms affect empathy for outgroup members by altering how much outgroup members' misfortunes are perceived as deserved. Thus, in this chapter, I will emphasise the specific aspects of the research design that are of relevance to the current research question.

Method

Participants. Eligible participants had to reside in the UK, have no formal education in psychology, and have no history of diagnosed mental health disorders or autism. A total of

352 participants were recruited through social media and university recruitment pools for a two-part experiment. Subsequently, 19 participants were excluded because they either did not accurately summarise outgroup intergroup empathic norms presented in their experimental condition or they failed two or more out of six attention checks embedded in the main task, leaving an analysable sample of 333 participants, 114 in the impartial outgroup empathy condition (40 men, 72 women, 2 preferred not to say; M_{AGE} = 29.54, SD_{AGE} = 10.66), 104 in the biased outgroup empathy condition (38 men, 66 women; M_{AGE} = 29.06, SD_{AGE} = 9.55) and 115 in the control condition (45 men, 68 women, 2 non-binary; M_{AGE} = 30.97, SD_{AGE} = 10.34). If participants failed only one attention check embedded in the main task, only the trial where they made the mistake was excluded ($N_{EXCLUDED}$ = 67).

Procedure. As was outlined under Experiment 6a, Experiment 6b was divided into two parts. Below, I primarily discuss the impartial and biased outgroup empathy conditions, while the control condition is described more briefly, with the purpose of highlighting its differences in comparison to the two experimental conditions where outgroup intergroup empathic norms were explicitly communicated.

First part of the experiment. Participants were randomly assigned to a minimal group (Square or Triangle) and asked to describe one negative personal experience. Participants were told that between the two parts of the experiment, collected personal experiences would be shown to randomly selected members of both minimal groups, to assess how people perceive strangers' experiences and how much empathy they tend to show for them. As pointed out in Experiment 6a, this was only a cover story allowing us to explain where the information about outgroup intergroup empathic norms came from in the second part of the experiment, and no participant in fact ever read these personal experiences. Furthermore, when describing their negative personal experiences, participants were instructed to do so in a way that corresponded to how the stimuli (i.e., short stories) used in the second part of the

experiment were structured. This was done to convince participants that the included stimuli really belonged to other people taking part in the study.

Second part of the experiment. After a week, the same group of participants was assigned to one of the experimental conditions: (a) impartial outgroup empathy condition, (b) biased outgroup empathy condition and (c) control condition. The first two conditions included explicit feedback about minimal outgroup members' intergroup empathy (see Figure 4.2). In the impartial outgroup empathy condition, participants were informed that randomly selected members of the minimal outgroup, who supposedly responded to personal experiences collected in the first part of the experiment, showed the same level of empathy for the participants' minimal ingroup as they did for their own minimal group (see Figure 4.2A). In other words, in the impartial outgroup empathy condition, minimal outgroup members were portrayed as not discriminating in their empathy based on the empathic targets' group membership. In contrast, in the biased outgroup empathy condition, participants were informed that randomly selected members of the minimal outgroup showed stronger empathy for members of their own minimal group than for participants' minimal ingroup, thus discriminating in their empathy based on empathic targets' group membership (see Figure 4.2B). Having received empathic feedback, participants in the impartial and biased outgroup empathy conditions were asked to summarise it before continuing to the main task. Participants in the control condition did not receive any empathic feedback, but instead immediately proceeded to the main task described below.

In the main task, included in all three experimental conditions, participants were presented with short stories – eight describing negative events supposedly reported by other participants belonging to their minimal ingroup, and eight describing negative events supposedly reported by other participants belonging to their minimal outgroup. The pre-selected short stories described experiences where the empathic target felt one of the

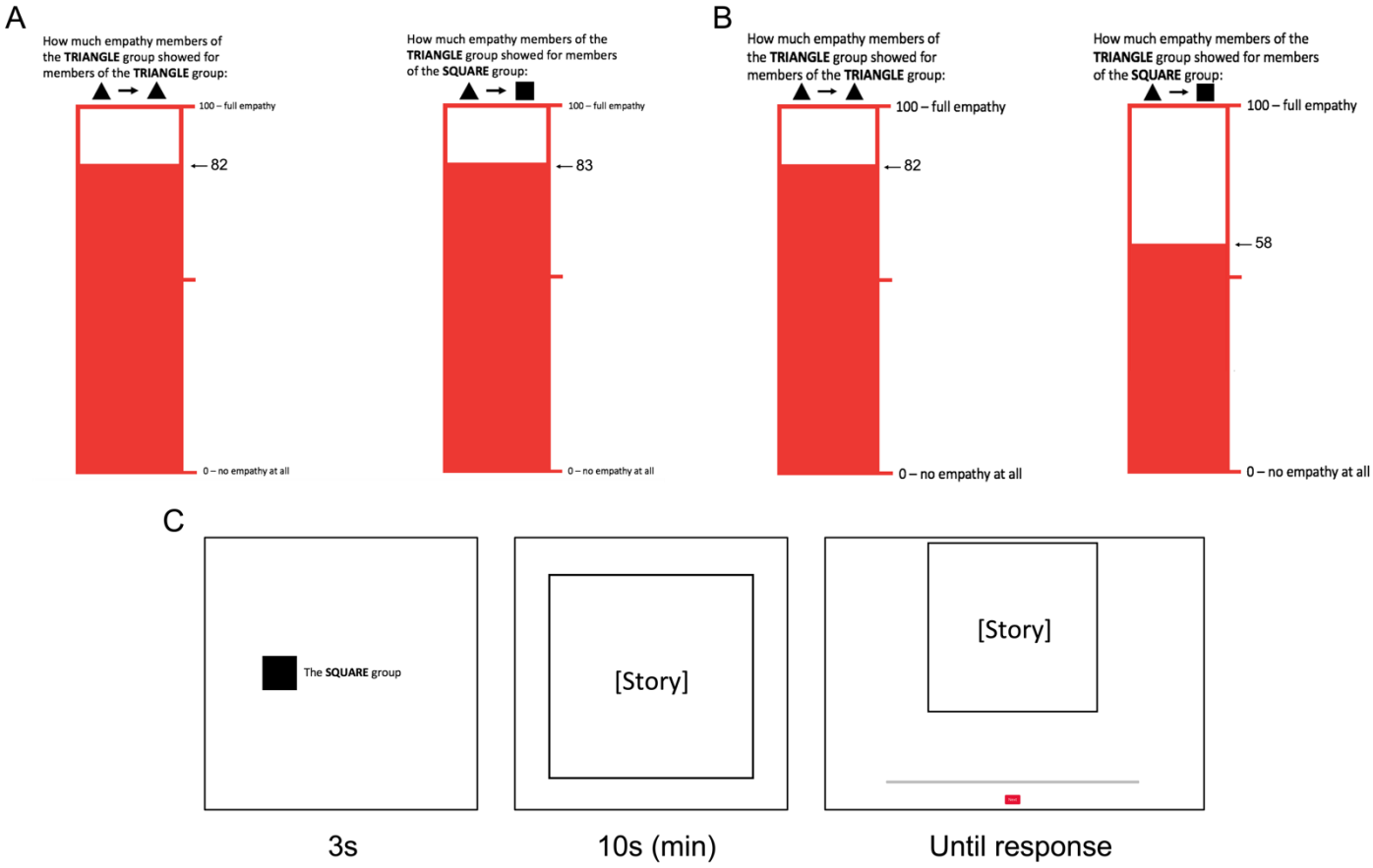
following emotions – anger/frustration (three empathic targets from each minimal group), sadness/disappointment (three empathic targets from each minimal group), and guilt (two empathic targets from each minimal group). Anger, guilt and sadness were chosen as they vary with respect to how much they trigger deservingness attributions (Feather, 2006; Pivetti et al., 2016; Smith & Lazarus, 1993). Furthermore, empathic targets' group membership was illustrated with a verbal cue at the beginning of each trial and with the corresponding shape in which participants read personal stories (e.g., participants read empathic targets' experiences in a triangle or a square box corresponding to the targets' minimal group membership, see Figure 4.2C). Each story had a specific part clearly marked at or close to the end signifying the peak of the described emotional experience and participants were asked to focus on the marked section when giving responses. Finally, no stories included affective statements, but rather just a factual description of the emotion-inducing event. After reading each story, participants gave emotion perception (*How bad/good do you think this person felt?*), empathy (*How bad does what happened to this person make you feel?*), schadenfreude (*How good does what happened to this person make you feel?*) and deservingness estimates (*How much do you think this person deserved what happened to them?*) in a randomised order.

Please note that the short stories included in the main task were not in fact produced by real participants but were selected based on a separate pilot study. The stories reflecting different emotions were matched for valence ($p = .736$), but the pilot study revealed baseline differences in perceived deservingness ($F(2, 38) = 44.10$, $p < .001$), with experiences characterised by guilt ($M = 54.20$, $SD = 27.91$) being perceived as most deserved, followed by experiences characterised by anger/frustration ($M = 30.08$, $SD = 25.39$), while experiences characterised by sadness/disappointment were perceived as least deserved ($M = 20.20$, $SD = 23.03$). As suggested in the previous paragraph, these pilot study results reflect natural variation between the three discrete emotions (Feather, 2006; Pivetti et al., 2016; Smith &

Lazarus, 1993). Thus, in further analyses and interpretations of results, I do not discuss baseline differences between emotions, as these are not relevant to the current research questions. Instead, I focus on main effects and interactions which include two factors of interest – (a) empathic targets' group membership (levels: ingroup member, outgroup member) and (c) experimental condition (levels: impartial outgroup empathic norms, biased outgroup empathic norms, control). More details on the pilot study and its results can be found in the Appendix 3.2, while all short stories are available on Open Science Framework under Experiment 3 (https://osf.io/d938s/?view_only=842a24066afe490db4bccda82d5b566d).

Figure 4.2

Empathic feedback and an example of a main task trial



Notes. (A) Outgroup intergroup empathic norms illustrated to participants who belonged to the "Square" group and who were assigned to the impartial outgroup empathy condition. (B) Outgroup intergroup empathic norms illustrated to participants who belonged to the "Square" group and who were assigned to the biased outgroup

empathy condition. (A-B) Prior to receiving any feedback about outgroup intergroup empathic norms, empathy was defined to participants as follows: "In psychology, empathy refers to situations when a person sees someone expressing an emotion and starts to feel a similar emotion. For instance, the person sees someone crying and feels sad with them, or the person hears about a friend's promotion and feels excited with them." (Depow et al., 2021). (C) Example trial from the main task in which the empathic target belonged to the "Square" group. The short story was always available on the screen while participants were giving responses.

After the main task, participants in the control condition were asked about their acontextualised perceptions of empathic norms, as was described in Experiment 6a. Finally, all participants completed a debrief sheet which included one question about the convincingness of the stories (*To what extent were you convinced that the stories you read really belonged to other research participants?*), included in all three experimental conditions, and one question about the convincingness of illustrated outgroup intergroup empathic norms (*To what extent were you convinced by the empathy feedback we provided you with at the start of the study?*), included only in impartial and biased outgroup empathy conditions. Both questions required responses on a seven-point scale from "not at all convinced" to "absolutely convinced".

Results

To examine whether deservingness, empathy, and schadenfreude responses in the main task varied by experimental condition (levels: impartial outgroup empathy, biased outgroup empathy, control), empathic targets' group category membership (levels: ingroup member, outgroup member) and dominant emotion (levels: anger, guilt, sadness), I fitted a series of linear mixed models with these three factors included as fixed effects. All fitted models included a random intercept for participants and a random intercept for stimuli (i.e., short stories).

All analyses yielded a single significant effect – a main effect of experimental condition on self-reported empathy ($F(2, 331.7) = 6.20$, $p = .002$) – but no significant effects

on either deservingness (all *p*s > .441)[63] or schadenfreude (all *p*s > 153)[64]. Pairwise comparisons revealed that participants in the impartial outgroup empathy condition reported more empathy (*M* = 60.58, *SD* = 27.51) for empathic targets regardless of their group category membership in comparison to participants in the control condition (*M* = 52.41, *SD* = 30.64, *z* = 3.50, *p* = .001), while the difference between participants in the biased outgroup empathy condition (*M* = 57.45, *SD* = 29.37) and participants in the control condition was not statistically significant (*z* = 2.05, *p* = .082). Finally, no significant differences in empathy were observed between participants in impartial and biased intergroup empathy conditions (*z* = 1.37, *p* = .171). While outgroup empathic norms did influence participants' self-reported empathy, this result was not in line with the starting hypothesis which specified that their effect should be outgroup-specific, influencing empathy for outgroup but not ingroup empathic targets in the main task. Finally, as I observed no effects of experimental condition on perceived deservingness of negative experiences, I did not proceed to test the proposed mediation model.

Focusing on the control condition, when perceptions of outgroup intergroup empathic norms – operationalised as a difference score between how much empathy participants expected minimal outgroup members to show for (a) minimal outgroup targets versus (b) minimal ingroup targets – were included in the model, they predicted participants' baseline levels of empathy regardless of empathic targets' group category membership (*F*(1, 113.61) = 5.94, *p* = .016), such that participants with more impartial perceptions of outgroup empathic norms reported higher empathy overall in the main task. However, this effect became non-

[63] As expected, I also observed a main effect of the dominant emotion on perceived deservingness (*F*(2, 13) = 18.34, *p* < .001) with pairwise comparisons revealing the same pattern of differences as in the pilot study. These results are reported for transparency and are not further discussed.

[64] While the primary focus of Experiment 6b was on deservingness, empathy, and schadenfreude, I also analysed whether experimental condition, empathic targets' group membership and dominant emotion predicted how participants judged empathic targets' emotions, but no significant effects were observed (all *p*s > .073).

significant once corresponding perceptions of ingroup intergroup empathic norms (r = .51, p < .001) were controlled for in the model (F(1, 112.44) = 2.70, p = .103). Similarly, perceptions of outgroup intergroup empathic norms did not predict participants' baseline deservingness responses (F(1, 113) = 0.56, p = .457), nor did they interact with other factors of interest (all ps > .384)

Convincingness of experimental conditions. While no differences were found between participants assigned to the three experimental conditions in terms of how much they believed the described emotional experiences belonged to actual people ($M_{IMPARTIAL}$ = 5.57, $SD_{IMPARTIAL}$ = 1.69; M_{BIASED} = 5.00, SD_{BIASED} = 1.86; $M_{CONTROL}$ = 5.21, $SD_{CONTROL}$ = 1.90; F(2, 328) = 2.75, p = .065), the impartial and biased outgroup empathy conditions differed in how convincing the corresponding experimental manipulation was. Specifically, the manipulation used in the impartial outgroup empathy condition (illustrated in Figure 4.2A) was perceived as more convincing (M = 5.47, SD = 1.66) than the manipulation used in the biased outgroup empathy condition (illustrated in Figure 4.2B, M = 4.02, SD = 2.01, t(215) = 5.82, p < .001). However, the extent to which experimental manipulations were perceived as convincing did not moderate any effects of experimental condition on empathy (all ps > .195), schadenfreude (all ps > .102) or deservingness (all ps > .516).

Discussion

Experiment 6b was designed to examine whether one path through which outgroup intergroup empathic norms influence own empathy for outgroup members involves changing how much outgroup misfortunes are subsequently perceived as deserved. The main hypothesis tested in this work was that forming a perception of impartial empathy's normativity among outgroup members should reduce attributions of deservingness after learning about outgroup misfortunes, resulting in increased empathy for outgroup members and correspondingly, reduced intergroup empathy bias. In contrast, learning that biased

empathy is normative among outgroup members was expected to lead to outgroup misfortunes being construed as more deserved, resulting in reduced empathy for outgroup members and exacerbated intergroup empathy bias. Experiment 6b did not provide evidence in support of this outgroup-centred hypothesis. Specifically, rather than encouraging empathy exclusively for outgroup empathic targets, impartial outgroup empathic norms were shown to enhance overall expressions of empathy only in comparison to the control group, regardless of whether the empathic target was an ingroup or an outgroup member. However, the results do not point to the conclusion that this general empathy boost stemmed from impartial empathic norms affecting deservingness appraisals. While these findings can broadly be placed in the wider literature showing that empathic norms, illustrated in contexts with no salient group divisions, influence people's motivation to empathise (e.g., Weisz et al., 2022), it is important to carefully interpret them considering the current paradigm and research design.

The first characteristic of Experiment 6b that might have played a role in shaping the results was that outgroup intergroup empathic norms were illustrated in relative terms – how much outgroup members empathised with participants' ingroup members *compared to* how much they empathised with their own group members – rather than absolute terms – how much outgroup members empathised with participants' ingroup members with *no reference point for comparison*. This meant that the experimental manipulation used in Experiment 6b effectively communicated two pieces of social information: (a) how much *intergroup* empathy was normative among outgroup members, and (b) how much *intragroup* empathy was normative among outgroup members. These could have independently influenced how much empathy participants in the impartial outgroup empathy condition showed for outgroup and ingroup empathic targets respectively compared to participants in the control condition, leading to a general as opposed to an outgroup-specific increase in empathy. Therefore, it is

possible that had the experimental manipulation communicated outgroup intergroup empathic norms in absolute terms, participants would have shown an increase in empathy for outgroup, but not ingroup empathic targets (see Kteily et al., 2016 and Livingstone et al., 2020 for examples of "absolute" experimental manipulations). However, this remains a conjecture, and future research should more closely compare how people process social information that is relevant to intergroup relations, when it is communicated in relative terms (e.g., how they act towards us *in comparison to* how they act towards their own group or even other social groups; what is our position in the social hierarchy *in comparison to* them) versus absolute terms (e.g., how they act towards us; what is our position in the social hierarchy).

Second, the experimental manipulation used to illustrate biased outgroup empathy was too strong for the context of minimal groups, which potentially diminished the manipulation's effect on participants' own reports of empathy. To get a sense of how far this experimental manipulation was from participants' pre-existing beliefs about intergroup empathy in minimal intergroup environments, it is useful to contrast the intergroup empathy bias participants in the control condition expected outgroup members to show (on average, 6.9 points on a 100-point scale) and the intergroup empathy bias portrayed in the manipulation (24 points on a 100-point scale). Other, more informal indicators of participants' surprise at and disapproval of outgroup's intergroup empathy bias could be found in their responses summarising the empathic feedback they had been exposed to (e.g., "Squares are selfish", "I am quite surprised", "This is irrational, groups are entirely arbitrary", "I am unsure why there was such a large difference"). Nevertheless, how convinced participants were by the experimental manipulations made no difference to how much empathy they reported in the main task. In other words, the fact that the biased outgroup empathy condition did not influence participants' own empathy as hypothesised cannot be explained away by the manipulation's unconvincingness. Having more convincing

experimental manipulations would have certainly been better, but in the case of Experiment 6b, it is more likely that the overall experimental set-up, wherein the minimal group paradigm was combined with very detailed and potentially individuating stimuli to study a very socially desirable process such as empathy, was suboptimal. For comparison, in most past research attempting to examine empathy in the context of minimal groups, stimuli used to assess it were very low-intensity and lacked personal details that would arguably stimulate participants' greater involvement (e.g., "Brandon accidentally walked into a glass door" was one of the stimuli used in Cikara et al., 2014), thus potentially leaving more room for minimal groups to exert their influence. By contrast, stimuli included in Experiment 6b were significantly longer (5-6 sentence paragraphs), more detailed, and of higher intensity (i.e., going as intense as to have an empathic target tell their story of being rejected from university), all of which could minimise the scope of influence empathic norms and/or minimal groups could have on participants' reports of empathy. Additionally, such detailed and personalised stimuli could have made participants focus on empathic targets' personal rather than on their minimal group identity (Bruneau et al., 2015; Sheng & Han, 2012). In summary, while recording participants' empathy for Brandon ("… accidentally walking into a glass door", Cikara et al., 2014) does not capture real-world empathy in all its intensity and nuance, if one is to rely on minimal groups as a proof-of-concept research environment, one may be better served by adjusting the experimental stimuli and overall set-up to account for the effect size usually associated with the minimal group paradigm.

Third, the short stories included in the main task were potentially inadequate to assess deservingness appraisals. Going back to the literature, spontaneous deservingness appraisals are theorised to arise from perceivers' evaluations of how targets' outcomes relate to actions that produced them (Feather, 2006). For instance, if a perceiver knows the target engaged in a positive action (e.g., a colleague worked hard on a project) which resulted in a positive

outcome (e.g., the colleague was rewarded for their hard work), they are likely to perceive the target's outcome as deserved and feel pleasure in response. On the other hand, if the same target engaged in a negative action (e.g., a colleague took credit for someone else's work) which resulted in a positive outcome (e.g., the colleague was again rewarded for their "hard work"), the perceiver is more likely to see the target's outcome as undeserved and feel resentment and gluckschmerz in response. While in these examples the target's action that preceded the outcome is known, rendering deservingness appraisals less ambiguous, action-related information was not consistently and unambiguously available in all included short stories. One of the stories, for example, described an empathic target who fell ill right before travelling and was forced to cancel the scheduled trip. The limitation of this story was that it included no information about the target's actions which came before them getting sick, which could have substantially disambiguated deservingness appraisals. If this short story were to be used as a stimulus again, participants could be told that the empathic target made an effort to avoid getting sick before travelling (which would more directly communicate an *undeserved* misfortune) or that the empathic target was careless about their health (which would more directly communicate a *deserved* misfortune). Without clear action-related information, participants probably created an idiosyncratic background explanation to assess how much this target's outcome (e.g., getting sick) was deserved. Thus, future research examining the role of deservingness in intergroup empathy should develop the experimental stimuli in a more principled manner, particularly with respect to transparency of targets' actions giving rise to outcomes (also see Feather & Mckee, 2014; Powell & Smith, 2013; van Dijk et al., 2008).

In summary, Experiment 6b sought to elucidate whether outgroup intergroup empathic norms affect own intergroup empathy by changing the extent to which outgroup's adversities are subsequently judged as deserved. Although Experiment 6b paved the way for

future research to address the underlying mechanisms whereby empathic norms exert their influence, in itself, it does not allow any strong conclusions about deservingness appraisals being one of these mechanisms, due to several identified methodological shortcomings. Thus, instead of abandoning deservingness as a candidate mechanism in response to the non-significant results, Experiment 6b should solely be approached as a stepping stone in this line of research, which lists a series of considerations future research needs to take into account and offers a set of methodological recommendations for those aiming to make use of the minimal group paradigm to address similar research questions.

Experiment 7

Imagine scrolling through social media to learn about recent political events in the US and stumbling upon two articles: (a) one describing a Republican voter suffering an unintentional firearm injury, and (b) one describing a Democrat voter whose home had been burgled. As your attention is captured by both headlines, you decide to go to the comments section and there you see the following: (a) in response to the first article, you see what seem to be Democrat voters' comments such as "oh, well, you can't expect us to care about someone who probably does not support gun control themselves", and (b) in response to the second article, you see what seem to be Republican voters' comments like "this is scary, nobody deserves to go through this regardless of what their political beliefs are". Depending on your political affiliation, how will your own empathy for a political outgroup member described in one of the articles be affected after seeing Democrat and Republican voters' apathetic and empathetic comments respectively?

Existing research has indeed documented that one of the most effective interventions designed to encourage intergroup empathy are those leveraging empathic norms (Gubler et al., 2015; Hein et al., 2015; Nadler & Liviatan, 2006; Nook et al., 2016; Saguy et al., 2015; Tarrant et al., 2009), thus providing ample evidence that intergroup empathy is not immune

to normative influence. However, the dominant focus on either the influence of ingroup or outgroup intergroup empathic norms taken separately rather than in combination makes it difficult to draw conclusions about how intergroup empathy operates when both intergroup empathic norms are salient, as in the starting example. To understand how people navigate this multi-norm space, Experiment 7 compares the levels of empathy people display[65] for outgroup members when ingroup intergroup empathic norms and outgroup intergroup empathic norms are aligned (i.e., both promote apathy or both promote empathy) and when they are misaligned (i.e., one set of norms promotes apathy, while the other promotes empathy). Experiment 7 thus addresses the following question: Do people prefer to *reciprocate* the outgroup's emotional responses to their ingroup's outcomes, even when their ingroup responds differently to outgroup's outcomes, or do they prefer to *conform* to ingroup's emotional responses, even when the outgroup responds differently to ingroup's outcomes? In other words, does *reciprocity* or *conformity* have a greater influence on intergroup empathy?

To date, very few studies have simultaneously examined reciprocity and conformity in the context of any social behaviour but the work of Romano and Balliet (2017) is a notable exception. Specifically, Romano and Balliet (2017) had participants play a task where they had to decide how to split their endowment (i.e., 100 fictitious tickets) with a partner. Importantly, before making the decision, participants learnt (a) how other members of the minimal ingroup interacted with the same partner in the past (i.e., cooperatively, or non-cooperatively) and (b) how this partner interacted with minimal ingroup members in the past (i.e., cooperatively, or non-cooperatively). These two pieces of information were manipulated such that in two experimental conditions, conformity and reciprocity directly competed for

[65] The experimental paradigm adopted in Experiment 7 is better suited to assess *displays* of intergroup empathy than *feelings* of intergroup empathy.

influence over participants' cooperative behaviour. In these conditions, if participants' decisions aligned with how their group members behaved, they were considered to have *conformed* to ingroup's behaviour, and if participants' decisions aligned with how their partner behaved, they were considered to have *reciprocated* their partner's behaviour. Across three experiments, participants were shown to both reciprocate and conform, with reciprocity exerting greater influence on cooperation than conformity and conformity being at least partly driven by participants' reputational concerns (i.e., concerns about how they would be evaluated by minimal ingroup members). Using the same set-up, Bică (2022) showed that people with high levels of social anxiety tended to conform more than reciprocate potentially due to their underlying motivation to seek approval and avoid disapproval from ingroup members. Both studies (Romano & Balliet, 2017; Bică, 2022) thus agree that individuals' reputational concerns in relation to their ingroup members meaningfully predict which source of normative influence they decide to follow. Expanding this work to the domain of intergroup empathic norms, Experiment 7 took a twofold approach to examining the role of reputational concerns.

Under the first approach, Experiment 7 investigated whether the influence of ingroup and/or outgroup intergroup empathic norms may differ between public settings, where one's reputation is at stake, and private settings, where one is away from the prying eyes of others. Comparing public and private settings is a well-established approach to examine how people's behaviour changes with enhanced reputational concerns and research relying on this approach has consistently documented that both children over the age of five and adults act more prosocially in the presence of others than when they are alone (Barclay, 2013; Bradley et al., 2018; Bereczkei et al., 2007; Engelmann et al., 2012; Engelmann & Rapp, 2018). While most past work has examined reputational concerns in the context of cooperation (e.g., Manrique et al., 2021; Roberts et al., 2021; Wu et al., 2016), empathy is also likely to activate

these concerns due to its connotations of morality (Fowler et al., 2021; Schumann et al., 2014). In essence, expressing empathy often leads to a positive reputation (Bieberstein et al., 2021; Fannes & Claeys, 2022; Schoofs et al., 2019) and for this reason, people might display empathy in the presence of others even when they do not genuinely feel it. To illustrate, imagine a situation where one's ingroup members (e.g., the UK public) overwhelmingly display empathy for members of another social group (e.g., Ukrainian refugees). In this situation, even if one (e.g., a single British citizen) does not feel empathy for the same social group, one may still be motivated to comply with these ingroup intergroup empathic norms in public to achieve a positive evaluation from others and/or to avoid the risk of facing disapproval and social exclusion (Cavazza et al., 2014). However, complying with ingroup intergroup empathic norms in public (e.g., in conversations) does not guarantee that one would similarly display intergroup empathy outside of the spotlight (e.g., while scrolling through the news in the privacy of their own home). The previous example describes a situation where people might be motivated to display intergroup empathy they do not genuinely feel to protect or enhance their reputation. However, it is important to recognise there are also circumstances in which intergroup empathy might be counter-normative, and thus not welcomed or even punished by others (e.g., a general reputational prescription to empathise with others might be incompatible with ingroup reputational prescription to apathetically ignore outgroup members, see Berman & Silver, 2022; Hasson et al., 2018; Kawamura & Kusumi, 2020; Porat et al., 2016; Raihani & Power, 2021 for related work).

In summary, recognising that reputational concerns are inherently implicated in the expression of empathy and in norm compliance more broadly, Experiment 7 examined how people's compliance to ingroup and/or outgroup intergroup empathic norms changes when their displays of intergroup empathy are publicly visible and thus subject to reputational scrutiny as opposed to private and thus protected from reputational scrutiny. Furthermore,

due to inclusion of experimental conditions with different sets of empathic norms, Experiment 7 also made it possible to examine how people's norm compliance varies across public and private settings depending on the specific content of empathic norms (i.e., whether they prescribe apathy or empathy across group lines) and its source (i.e., ingroup or outgroup members). Finally, some authors have advised against equating norm compliance in private with psychological acceptance of the norm (i.e., private feelings of intergroup empathy, Zaki et al., 2011). In line with this recommendation, I interpret participants' responses given both in public and in private as *displays* of intergroup empathy rather than internal *feelings* of intergroup empathy. Nevertheless, comparing public and private settings is still informative regarding the role of reputational concerns.

Under the second approach, I followed the example of Romano and Balliet (2017) and included a self-report measure of reputational concern, wherein participants were asked to rate their level of concern about ingroup members' evaluation of them (example item: *It's important to me that my group members have a positive evaluation about me.*). This measure was then included in analyses to examine whether individual differences in reputational concern predict individual differences in empathic norm compliance. Of note, most past work has emphasised that reputational concerns are primarily felt in relation to ingroup members (Everett et al., 2015; Yamagishi & Kiyonari, 2000) unless the underlying coalitional structure is misaligned with group distinctions (i.e., participants are interdependent with outgroup members, Romano et al., 2017). Given that Experiment 7 fully aligned group distinctions with coalitional structure (i.e., participants were interdependent only with ingroup members and did not expect any future interactions with outgroup members), reputational concerns were only measured in relation to ingroup members.

Finally, Experiment 7 included another component, wherein it touched on the question of how perceptions of intergroup empathic norms come to be overly negative and

inaccurate, as was documented in previous research (e.g., Penić et al., 2021) as well as in Chapter 3. It is first important to recognise that intergroup empathic norms can be communicated in various ways, and correspondingly, perceptions of intergroup empathic norms can be acquired in various ways. For example, you might start to perceive intergroup empathy as counter-normative because an outgroup member you approached offered you no help when you expressed distress (i.e., learning from direct experience), because you overheard an outgroup member make an insensitive remark about an ingroup member's experience (i.e., learning through observation) or because you came across a scientific article suggesting that members of one social group do not show much empathy for members of a different social group (i.e., learning through instruction, Olsson & Phelps, 2007).

In contrast to the real-world variability in communication of intergroup empathic norms and acquisition of intergroup empathic norm perceptions, most research tracking the influence of intergroup empathic norms on personal intergroup empathy has been homogenous in its approach to how intergroup empathic norms are conveyed. Specifically, most research to date has used explicit instruction, usually in the form of one-off summary information (e.g., previous participants' empathy ratings – Nook et al., 2016; verbal summary that participants' ingroup tends to show "high levels of compassion, tenderness, and empathy towards others" – Tarrant et al., 2009; article about Israeli doctors who volunteered to treat Palestinian children – Saguy et al., 2015, see Tankard & Paluck, 2016) and has made little attempt to unravel the learning mechanisms contributing to the formation and updating of empathic norm perceptions in intergroup environments (but see Hein et al., 2016). Although there may be some circumstances in which intergroup empathic norms are communicated in such a summarised, instruction-like manner, what often happens in the real world is that people observe emotions and behaviours of individual group members over time (Hertz, 2021), and derive a group-level norm from these observations (Hackel et al., 2022). Thus, in

Experiment 7, I attempted to mimic this learning of intergroup empathic norms through observation. To achieve this, I designed an empathy learning task where participants witnessed how individual ingroup and outgroup members emotionally responded to each other's positive or negative outcomes. This task allowed examination of how participants learn about operative intergroup empathic norms from individual observations of group members' emotions and how they adapt to these norms in their own empathic responding.

Going back to the question of how perceptions of intergroup empathic norms come to be overly negative and inaccurate, one intuitive explanation immediately comes to mind. In particular, negative perceptions of intergroup empathic norms might simply be a product of disproportionate exposure to negative intergroup attitudes and interactions, thus providing an accurate reflection of the unrepresentative section of reality that group members experience. For instance, while social media provide an ostensibly unbounded space for intergroup interactions, especially among groups that are otherwise segregated (Brown & Enos, 2021; van Ham & Tammaru, 2016), they also contribute to the spread of animosity and divisiveness, with negative content about outgroups being particularly likely to go "viral" (Rathje et al., 2021; Stella et al., 2018; Van Bavel et al., 2021). Consequentially, such social media experience can distort people's perceptions of intergroup emotions and intergroup relations more broadly (Moore-Berg et al., 2020; Ruggeri et al., 2021). In summary, this exposure-centred interpretation assumes that biases in perceptions of intergroup empathic norms stem from people's unbiased learning of unrepresentative content.

While nobody would deny that people's perceptions depend on their social surroundings, laying the blame for overly negative perceptions of intergroup empathic norms solely on negative exposure is largely inconsistent with previous work. First, as documented in Chapter 3, people amplify intergroup empathy bias even in minimal intergroup environments, where no prior exposure exists. This amplified intergroup empathy bias is thus

better understood as a reflection of people's intuitive beliefs about how groups affect emotions. In other words, in the absence of any exposure that would create opportunities for learning about minimal groups and their relations, people arguably endorse an intuitive social heuristic that others empathise more with own-group as opposed to other-group members. As such, this minimal group result suggests that asymmetrical exposure to depictions of intergroup negativity is not the root cause of biases in perceptions of intergroup empathic norms, but instead more likely an amplifier of them, with the extent of amplification being proportional to the extent to which the content received by people deviates from how most group members feel. Second, the exposure-centred explanation fails to take into account the substantial literature showing that people approach the social world in a confirmatory manner through the lens of their pre-existing beliefs (e.g., Edwards & Smith, 1996; Lord et al., 1979; Leong & Zaki, 2017). In intergroup environments, the influence of pre-existing beliefs can reverberate in different ways – for example, people prefer content that affirms their stereotypical views (Allen et al., 2009), avoid the entire outgroup after initial negative encounters with individual outgroup members (Allidina & Cunningham, 2021) or when avoidance is not an option, as is the case in Experiment 7, pre-existing beliefs have been shown to shape how people experience intergroup interactions and what social representations they form on the basis of them (Stillerman et al., 2020). Taken together, these findings suggest that biases in perceptions of intergroup empathic norms might arise as a product of biased learning, even when the content being learnt is not unrepresentative.

Given that Experiment 7 relied on the minimal group paradigm, it had full control over the content (i.e., intergroup interactions) participants were exposed to, making it well placed to assess the extent to which participants' final perceptions of intergroup empathic norms deviated from true intergroup empathic norms (i.e., intergroup empathy shown by ingroup and outgroup members on average). If participants' perceptions indeed deviate from

reality, further investigation of potential biases in observational learning is warranted, with three biases being particularly likely to contribute to discrepancies between perceptions of intergroup empathic norms and actual empathic norms: (a) people's biased pre-existing beliefs about intergroup empathy, (b) people's biased updating of perceptions of intergroup empathic norms such that these perceptions are updated more (or less) rapidly from negative "surprises" (i.e., situations where ingroup or outgroup members display less intergroup empathy than they were expected to display) than positive "surprises" (i.e., situations where ingroup or outgroup members display more intergroup empathy than they were expected to display, Müller-Pinzler et al., 2019; Sharot & Garrett, 2016), and (c) people's tendency to assume the tit-for-tat nature of intergroup relations wherein whatever one group does is automatically expected to be reciprocated by the other group and is thus generalised to the other group (i.e., reciprocity bias; Kamans et al., 2009; Moore-Berg et al., 2020). More details on how each of these biases might come into play in the context of the empathy learning task can be found under H4 in the "Hypotheses" section.

In summary, Experiment 7 addressed three research questions: (a) whether people preferentially conform to ingroup or reciprocate outgroup intergroup empathic norms, (b) whether individually or contextually determined reputational concerns predict which empathic norms people preferentially adhere to, and finally, (c) whether perceptions of empathic norms people form accurately reflect empathic norms they were exposed to and if not, whether this discrepancy is a consequence of observational learning being affected by biased pre-existing beliefs about intergroup empathy, biased updating and/or reciprocity bias. Before discussing methods and results, I will list the relevant hypotheses.

Hypotheses

In this section, I will briefly outline hypotheses pertaining to each research question. A separate hypothesis will be offered for each of the two approaches adopted to examine the

role of reputational concern. Hypotheses will occasionally be repeated when presenting the results, where deemed relevant.

H1. Participants will both conform to ingroup intergroup empathic norms and reciprocate outgroup intergroup empathic norms, with reciprocity having a greater influence on participants' intergroup empathy than conformity. This hypothesis assumes that intergroup empathic norms will influence participants' *intergroup empathy* (i.e., empathy for outgroup members) but not their *intragroup empathy* (i.e., empathy for ingroup members).

H2. I do not expect to observe any differences between participants' public and private expressions of intergroup empathy. Given that in Experiment 7 private expressions of intergroup empathy always come after extensive public expressions of intergroup empathy, it is possible that participants form a habit of emotionally responding in a certain way to outgroup members' outcomes (Wood & Rünger, 2016) and that this habit remains largely unchanged after participants are given the opportunity to respond anonymously, in the absence of reputational concerns[66]. Similarly, participants' expressions of intragroup empathy will remain unchanged across public and private settings.

However, looking at incongruent experimental conditions where conformity and reciprocity compete for influence, if there is one response that is likely to be changed by eliminating reputational concerns, it is participants' conformity to ingroup members' intergroup empathy. To better understand this possibility, it is useful to remember that participants in Experiment 7 are led to believe that they would only interact with ingroup

[66] Even though the first hypothesis was inspired by research on habitual behaviour, it is important to emphasise that Experiment 7 was not designed to investigate whether habit-like processes contribute to intergroup empathic norm compliance. As such, Experiment 7 should not be used to either support or discard the role of habits in intergroup emotions and behaviour but should instead be used to inspire future research in this direction.

members in the future and that their personal outcomes would depend on ingroup members' actions on this future occasion. Thus, it is reasonable to assume that this experimental set-up nudges participants to be concerned about how they are evaluated by ingroup members. If conformity to ingroup members' intergroup empathy is primarily driven by these ingroup-oriented reputational concerns, conformity might be reduced in private as anonymity eliminates these reputational concerns. In other words, one can imagine a situation where participants are motivated by reputational concern to comply with ingroup members' intergroup empathy in public, but they still privately think that these emotional reactions are inappropriate especially considering outgroup members' intergroup apathy. If this is the case, participants might lose some or all of their motivation to conform to ingroup members' intergroup empathy once reputational concerns are eliminated and might adjust their displays of intergroup empathy to reflect the emotional responses they genuinely think are appropriate. Given there is no reason to assume that participants' reciprocity of outgroup members' intergroup empathy similarly reflects reputational concerns, reciprocity is more likely to remain unchanged across private and public settings.

H3. When displaying intergroup empathy publicly, participants who report higher reputational concern will preferentially adhere to ingroup intergroup empathic norms over outgroup intergroup empathic norms while participants with lower reputational concern will preferentially adhere to outgroup intergroup empathic norms over ingroup intergroup empathic norms. In line with past work (see Romano & Balliet, 2017), this hypothesis only concerns incongruent experimental conditions, where the influence of ingroup intergroup empathic norms (conformity) and outgroup intergroup empathic norms (reciprocity) can be distinguished.

H4. Participants' perceptions of empathic norms will deviate from true empathic norms (i.e., average intergroup empathy shown by ingroup and outgroup members) such that

participants will think that both ingroup and outgroup members displayed less intergroup empathy than they in fact did. This underestimation of intergroup empathy will stem from one of the following biases in observational learning or their combination: (a) participants' negative pre-existing beliefs about intergroup empathy (i.e., negative priors about intergroup empathy), (b) participants' tendency to show increased updating of their intergroup empathic norm perceptions when intergroup empathy was lower than expected compared to when intergroup empathy was higher than expected (i.e., negativity bias) and (c) participants' tendency to generalise one group's emotions and actions to the other group (i.e., reciprocity bias).

First, a participant might start the empathy learning task with a pre-existing belief that people do not show much empathy or even show apathy for those with whom they do not share group category membership. Consequentially, if this participant is assigned to the experimental condition where ingroup members are empathetic and outgroup members are apathetic, this prior might slow down learning of disconfirming ingroup empathic norms and facilitate learning of confirming outgroup empathic norms on initial trials of the task, but its influence should ultimately be overwhelmed by the emotional reactions the participant is exposed to (Stillerman et al., 2020). In other words, if this participant only holds a negative prior about intergroup empathy in the absence of additional biases (i.e., negativity or reciprocity bias), this prior will affect initial learning but will soon lose influence in the light of consistent evidence.

Second, in addition to having a negative prior about intergroup empathy, this participant could display a negativity bias in their observational learning, wherein they update their perceptions of empathic norms more based on emotional reactions indicative of lower intergroup empathy than emotional reactions indicative of higher intergroup empathy. For instance, after a few initial trials, this participant has started to recognise that ingroup

members tend to display more intergroup empathy than outgroup members. To illustrate how negativity bias might come into play, I will discuss two hypothetical trials. On one trial, this participant sees an ingroup member showing lower levels of intergroup empathy than usual, and on another trial, they see an ingroup member showing higher levels of intergroup empathy than usual. In the first scenario, this participant might be likely to overweight a single display of lower intergroup empathy and to preferentially incorporate it in their perceptions of ingroup intergroup empathic norms. In the second scenario, this participant might be likely to underweight or "explain away" a single display of higher intergroup empathy, resulting in its reduced incorporation in participant's perceptions of intergroup empathic norms.

Third, this participant could also display a reciprocity bias in their observational learning such that observing emotions of one group does not only feed into perceptions of that group's intergroup empathic norms, but also into perceptions of the other group's intergroup empathic norms. To illustrate, imagine that the participant saw an outgroup member display intergroup apathy towards the ingroup. This display of intergroup apathy from the outgroup does not only inform participant's perceptions of outgroup intergroup empathic norms, but also of ingroup intergroup empathic norms as the participant can reasonably assume that ingroup members would reciprocate the apathy they received. Finally, although I focused on a single participant to offer illustrative examples of how different biases might characterise observational learning, it is important to note that these parameters are estimated on a participant-by-participant basis, thus allowing different participants to have different pre-existing beliefs, as well as negativity (or even positivity) and reciprocity biases of varying strength.

Method

Overview. To recap, Experiment 7 sought to test (a) whether people preferentially conform to ingroup or reciprocate outgroup intergroup empathic norms, (b) whether reputational concerns can help explain which empathic norms people adhere to, and (c) whether people's final perceptions of empathic norms accurately reflect the empathic norms they were in fact exposed to. These questions were addressed in the context of the empathy learning task which mimicked observational learning of intergroup empathic norms by placing participants in a social environment where they witnessed individual ingroup and outgroup members respond apathetically or empathetically to each other's outcomes. Specifically, participants were allocated to one of the four experimental conditions, where ingroup and outgroup intergroup empathic norms were either aligned (e.g., both promoted apathy or empathy for outgroup members) or misaligned (e.g., one promoted apathy, the other one empathy for outgroup members). As such, Experiment 7 was able to assess how participants' displays of intergroup empathy change as they get acquainted with operative intergroup empathic norms. To examine the role of reputational concerns, empathy learning task was divided into two parts – learning and test phase – where participants' displays of intra- and intergroup empathy were public and private respectively. This allowed me to assess how participants' displays of empathy for ingroup and outgroup members change when reputational concerns, associated with the initial presence of others, are eliminated. Finally, given that Experiment 7 had full control over the exact intra- and intergroup empathic norms participants were exposed to, it was well-placed to assess how participants' final perceptions of these norms, reported after completing the empathy learning task, deviate from reality. More details can be found in the appropriate sections below.

Participants and design. Eligible participants had to be nationals of United States, have English as their first language and have no formal education in psychology. Prolific

(www.prolific.co) was used to recruit 260 participants, out of which 9 participants were subsequently excluded because they failed at least one of the two attention checks, leaving a final sample of 251 participants (123 men, 124 women, 4 non-binary; M_{AGE} = 37.37, SD_{AGE} = 13.57). Experiment 7 used a 2 (ingroup intergroup empathic norms: apathy vs. empathy) x 2 (outgroup intergroup empathic norms: apathy vs. empathy) between-subjects design, with four experimental conditions which differed in the exact combination of intergroup empathic norms participants were exposed to in the main empathy learning task. Specifically, participants were assigned to one of the following experimental conditions: (a) ingroup empathy for the outgroup, outgroup empathy for the ingroup (i.e., bidirectional intergroup empathy condition, N = 62), (b) ingroup empathy for the outgroup, outgroup apathy for the ingroup (i.e., conform-to-empathise condition, N = 61), (c) ingroup apathy for the outgroup, outgroup empathy for the ingroup (i.e., reciprocate-to-empathise condition, N = 63), and (d) ingroup apathy for the outgroup, outgroup apathy for the ingroup (i.e., bidirectional intergroup apathy condition, N = 65). When determining the sample size, I followed an example of previous research using a similar experimental approach (Hertz, 2021) and aimed for a minimum of 60 participants in each of the four experimental conditions after exclusions. Finally, I ran post hoc sensitivity simulations for linear mixed models (1000 iterations, Green & MacLeod, 2016) which revealed that the observed linear mixed model estimates were, on average, 18.89% larger than the minimal detectable estimate with 80% power (median = 8.54%; range = [2.99%, 36.67%]).

Procedure. Participants were first randomly assigned to one of the two minimal groups – Lions or Tigers – and informed that they were to interact with other participants, also members of the Lions and the Tigers, in two tasks, both of which offered an opportunity to secure bonus payment. Participants were further told that in the first task, they would see both minimal ingroup and outgroup members, but the task outcomes would only reflect their

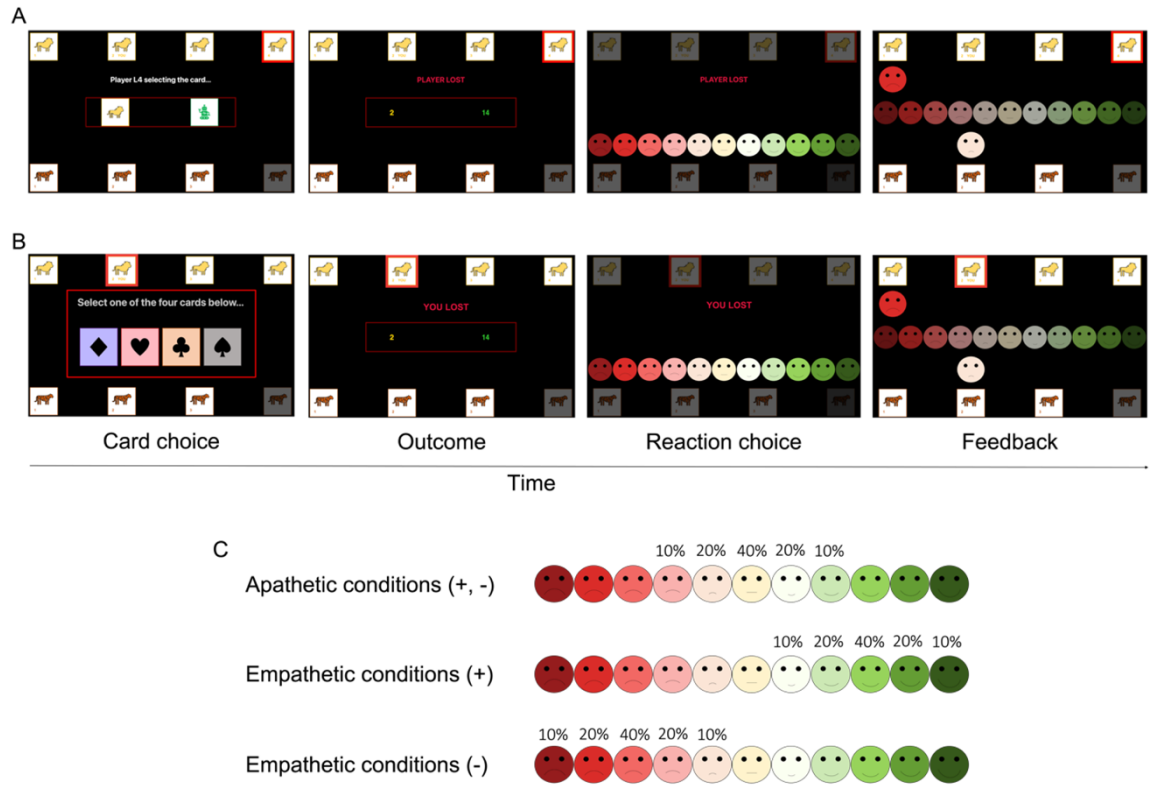
individual performance and would thus be independent from other participants (i.e., both minimal ingroup and outgroup members). The second task, however, was to be completed only with minimal ingroup members, and in this task participants' outcomes would directly depend on other minimal ingroup members' actions. In reality, the experiment contained a single task (empathy learning task), while the second task was only mentioned in the instructions to induce expectations of within-group interdependence and conceal the main reason behind group division. I decided to make within-group interdependence explicit as opposed to having participants assume it (Balliet et al., 2014; Brewer, 2008; Yamagishi & Kiyonari, 2000) so that any results could be interpreted in the context of known relational dynamics. Following the group allocation and the general introduction, participants proceeded to the empathy learning task, which was divided into two parts – learning (public) and test (private) phase.

Empathy learning task (learning phase). The empathy learning task was framed as a multiplayer card game that participants supposedly joined with six simultaneously active participants – three members of the minimal ingroup and three members of the minimal outgroup. On each trial, one active player would first choose one of the four cards, each associated with a numerical value in the background, after which the outcome of the trial would become known – (a) the active player lost if their chosen card gave a lower value than the value generated by the pre-programmed robot avatar, or (b) the active player won if their chosen card gave a higher value than the value generated by the pre-programmed robot avatar (sections "card choice" and "outcome" in Figures 4.3A and 4.3B). Outcomes presented on the screen were in fact unrelated to players' card choices, but instead, participants had an equal probability of witnessing and experiencing a loss or a victory over the course of the empathy learning task. Once the active player's outcome was known, participants were asked to report how they felt in response ("How bad/good do you feel because of this player's

outcome?" when the active player was someone else rather than the participant; "How bad/good do you feel because of your outcome?" when the active player was the participant) on an eleven-point emoji scale (section "reaction choice" in Figures 4.3A and 4.3B). Finally, after selecting an emoji, participants saw emojis supposedly chosen by one minimal ingroup member and one minimal outgroup member in response to the same active player's outcome (section "feedback" in Figures 4.3A and 4.3B). The exact emojis participants saw were pre-programmed and varied depending on the experimental condition with different emojis having a certain probability of being displayed (see Figure 4.3C). Participants completed 63 learning trials in total, out of which they were the active player on 9 trials (Figure 4.3B), while on other 54 trials they observed either a minimal ingroup member or a minimal outgroup member securing a positive or a negative outcome (Figure 4.3A).

Figure 4.3

Learning phase of the empathy learning task



Notes. (A) In this example trial, the active player was the fourth member of the Lions (L4), as signalled by the text and the red square surrounding the player's image, who suffered a loss. When the participant was not the active player, as in this example trial, they were led to believe that their responses on the emoji scale were public and visible to one ingroup member and one outgroup member on each trial. (B) In this example trial, the active player was the participant (i.e., the second member of the Lions, L2). The participant only saw the four cards when they were the active player. Furthermore, when the participant was the active player, as in this example trial, they were led to believe that their responses on the emoji scale were private. (C) Emotional reactions participants could see in different experimental conditions with their respective probabilities of appearing in the task. To illustrate, if a participant was assigned to the conform-to-empathise condition, this meant that at the feedback stage, they would see ingroup members displaying outcome-dependent empathy for outgroup members and outgroup members displaying outcome-independent apathy. Finally, plus (+) stands for positive outcomes, and minus (-) stands for negative outcomes.

A few characteristics of the learning phase are particularly important to emphasise. First, at the beginning of the task, participants joined a waiting room and over the course of two minutes, six other players ostensibly joined, resulting in a minimal ingroup that had four members with the participant, and a minimal outgroup that was supposed to have four

members, but one outgroup member never joined, and the task started without them. The two groups were different in size to provide an equal number of ingroup, and outgroup members that participants could observe and learn from, preventing them from generalising individual outgroup members' reactions more because there were more outgroup members. Second, the two minimal groups were not competing in the empathy learning task. Instead, active players always played the card game against a third party – a pre-programmed robot avatar – allowing me to interpret chosen emotional reactions as people's public displays of empathy rather than direct reactions to their ingroup's outcomes. Third, to justify the existence of empathic feedback, participants were led to believe their emotional reactions to other players' outcomes were also public and visible to one ingroup member and one outgroup member on each trial. On the 9 trials where they were the active player, participants' responses were private. These private responses were subsequently excluded from analyses as they reflected participants' emotional reactions to their own outcomes, and not empathy for other players' outcomes. Correspondingly, emotional reactions of active players were never displayed to participants. For instance, if player L3 (i.e., the third member of the Lions) was the active player on a certain trial, participants would never see L3's emotional reaction on this trial because this emotional reaction is not diagnostic of empathy. Fourth, while the emotional reactions ingroup and outgroup members displayed in response to each other's outcomes varied by experimental condition, intragroup emotional reactions were held constant across conditions and reflected outcome-dependent empathy (e.g., ingroup members would display positive emotional reactions when another ingroup member's outcome was positive or negative emotional reactions when another ingroup member's outcome was negative). Finally, even though the task operationalised intergroup empathic norms as behavioural regularities in the absence of explicit normative force (e.g., others' visible approval or disapproval), rendering them methodologically more similar to conventions than social norms

(Lewis, 1969; Nord, 1991), a theoretical assumption is that participants spontaneously imbue these behavioural regularities with normative force, viewing them as social norms or in other words, as standards of appropriate behaviour in the intergroup context they found themselves in.

Empathy learning task (test phase). After completing the learning phase, participants took a two-minute break and proceeded to the test phase, which did not include empathic feedback at the end of each trial. Specifically, participants again observed active players suffering a loss or securing a victory and were again asked to choose an emotional reaction in response, but they no longer had the opportunity to see how others in the same environment responded, which removed the possibility of further learning about empathic norms. Simultaneously, this rendered participants' own emotional reactions private, allowing me to examine whether compliance with intergroup empathic norms deteriorates after reputational concerns, stemming from the public nature of participants' emotional reactions in the learning phase, were removed[67]. Participants completed 21 trials in the test phase, out of

[67] Of note, one could argue that there might have been better ways to examine how participants' empathic norm compliance changes when reputational concerns are enhanced versus attenuated. For example, I could have (a) included separate public and private trials in the learning phase, (b) had participants provide both public and private responses on each trial of the learning phase, or (c) treated public and private responses as levels of a between-subject factor either from the beginning or after the learning phase. While considered, these alternative set-ups were ultimately not adopted for multiple reasons. First, interweaving public and private trials would significantly prolong the empathy learning task, lead to "gaps" in the learning models as the two types of trials cannot be combined and nudge participants to explicitly consider the authenticity of others' empathic displays. To elaborate on the last and the most critical reason, given that participants could track how their own responses changed across public and private settings, they could have started to wonder about the extent to which others' public responses in the task are authentic, which could have introduced unwanted noise to their learning of intergroup empathic norms (e.g., participants could have started to underweight displays of intergroup empathy thinking they are more likely to be "fake"). The same limitation applied to the option of having participants give both public and private responses on all trials. Finally, I decided against the option of splitting public and private responses into separate conditions because this would result in eight between-subject experimental

which they were an active player on 3 trials, while on other 18 trials they observed either a minimal ingroup member or a minimal outgroup member securing a positive or a negative outcome. The code to the full empathy learning task can be found here: https://github.com/jelkastojanov/js-empathy-learning-task.

Individual-differences measures. After the empathy learning task, participants completed seven additional measures in randomised order.

Reputational concern. As described in the introduction, I administered a measure of reputational concern (Romano & Balliet, 2017) to test whether participants with higher reputational concern preferentially adhered to ingroup over outgroup intergroup empathic norms in the learning phase while participants with lower reputational concern preferentially adhered to outgroup over ingroup intergroup empathic norms. In this measure, participants responded to four items (e.g., *It's important to me that my group members accept me*; α = 0.87) on a five-point Likert scale (from "totally disagree" to "totally agree").

***Group identification*.** Participants also responded to three items asking about their level of identification with both minimal groups (e.g., *I like members of the [minimal ingroup/outgroup]*; $\alpha_{INGROUP}$ = 0.88; $\alpha_{OUTGROUP}$ = 0.90; Vollberg & Cikara, 2022) on a five-point Likert scale (from "totally disagree" to "totally agree"). From these items, I originally planned to calculate an ingroup identification score as a difference between how participants responded to three ingroup-related items and three outgroup-related items. I then planned to examine whether participants with higher levels of ingroup identification aligned with ingroup intergroup empathic norms more than participants with lower levels of ingroup identification. However, given that participants filled in the group identification measure after completing the empathy learning task, I noticed that their identification responses varied as a

conditions (two levels of ingroup intergroup empathic norms x two levels of outgroup intergroup empathic norms x two levels of privacy) and would thus require an unfeasibly large sample size.

function of intergroup empathic norms they were exposed to and could thus not be treated as a pre-existing moderating variable. For this reason, group identification was not tested as a moderator, but I instead examined how intergroup empathic norms affected participants' level of identification with both minimal groups.

Learning check questions. Furthermore, participants completed learning check questions where they were first presented with different players' card game outcomes and then asked to report how members of their minimal ingroup, and the minimal outgroup most often emotionally responded to these outcomes in the empathy learning task. Participants gave their responses on the same 11-point emoji scale used in the task. These questions correspond to final perceptions of empathic norms participants formed in each experimental condition, thus allowing me to assess potential discrepancies between participants' perceptions and the empathic norms participants were in fact exposed to.

Other exploratory and control measures. I also collected and analysed the following measures. First, participants completed measures of dispositional empathic concern and perspective taking (Davis, 1973), which were used to detect any baseline differences between participants in the four experimental conditions and control for them accordingly. Analyses revealed that there were no differences in participants' baseline empathic concern across conditions (all *ps* > .104), but I found differences in baseline perspective taking between the experimental conditions where the ingroup was apathetic towards the outgroup (M = 27.70) and the experimental conditions where the ingroup was empathetic towards the outgroup (M = 25.74, p = .004). However, results remained the same regardless of whether perspective taking was statistically controlled for.

Second, participants responded to two items assessing the authenticity of ingroup and outgroup members' emotional reactions respectively (r_{INGROUP} = 0.59; r_{OUTGROUP} = 0.66, Côté et al., 2013) using a five-point Likert scale (from "completely disagree" to "completely

agree"). For example, one item read: "In the previous card game, [members of the minimal ingroup/outgroup] displayed feelings they did not really feel inside" (reverse scored item).

Third, I adapted the questions from Goldenberg et al. (2014) to assess participants' perceptions of the appropriateness of the emotional reactions displayed by minimal ingroup and outgroup members towards different empathic targets in the learning phase (e.g., *Do you agree that members of your group responded appropriately to other group members' outcomes in the card game?*; six-point Likert scale from "I don't agree at all" to "I absolutely agree"). In total, participants responded to four equivalent questions assessing how they perceived the appropriateness of both ingroup and outgroup members' intra- and intergroup emotional reactions in the task, but their assessments of intergroup emotional reactions' appropriateness were of primary interest in this study.

Finally, to examine whether observing others' emotional reactions in the task influenced personal feelings and inferences about group relations, participants reported their own feelings (e.g., *How do you feel towards the [minimal ingroup/outgroup]?*) and judged the feelings of average ingroup and outgroup members towards each minimal group (e.g., *How do you think an average member of the [minimal ingroup/outgroup] feels towards the [minimal ingroup/outgroup]?*). Participants gave their responses on multiple feeling thermometer scales ranging from 0 (*I/They have very negative feelings*) to 100 (*I/They have very positive feelings*, Converse et al., 1980; Haddock et al., 1993).

Analysis strategy. I addressed all hypotheses apart from the learning-related hypothesis using linear mixed models implemented in R (version 4.1.3) package lme4 (version 1.1.31, Bates et al., 2014), and lmerTest (version 3.1.3, Kuznetsova et al., 2017). All pairwise comparisons reported in the "Results" section used the Holm method for p-value correction from the package lsmeans (version 2.30.0, Lenth, 2016). In order for the dependent variable to reflect empathy, participants' emoji responses, originally collected on a

1 (i.e., most negative emoji) to 11 (i.e., most positive emoji) scale, were recoded to a -5 to 5 scale where -5 corresponded to the strongest counter-empathy (i.e., the most negative emoji to a positive outcome or the most positive emoji to a negative outcome), 0 corresponded to apathy (i.e., straight face emoji) and 5 corresponded to the strongest empathy (i.e., the most negative emoji to a negative outcome or the most positive emoji to a positive outcome). Regarding the models' random structure, I always started by fitting maximal models (Barr et al., 2013), and from there I simplified the random structure using Principal Component Analysis to determine the most relevant random components (Bates et al., 2015). Finally, participants' responses during the empathy learning task were modelled using an adapted Rescorla-Wagner learning model (Lockwood & Klein-Flügge, 2021; Rescorla & Wagner, 1972; Siegel & Allan, 1996; Sutton & Barto, 1998), which assumes that participants update their responses based on prediction errors, that is, the difference between an expected and an actual outcome (i.e., expected empathic feedback and actual empathic feedback in the context of the empathy learning task). Different models were created to reflect biased priors about intergroup empathy, biased updating, reciprocity bias or some combination of these biases. Model recovery, fitting and comparison were run in Python (version 3.9.7) with Jupyter Notebook (version 3.2.1). More details about specific linear mixed models, learning models as well as model fitting and recovery steps are provided at the beginning of the relevant sections below.

Results

Conformity versus reciprocity. To assess how conformity and reciprocity influence participants' empathy, I regressed participants' empathic responses onto fixed effects for empathic target (levels: ingroup member, outgroup member), ingroup intergroup empathic norm (levels: apathy, empathy) and outgroup intergroup empathic norm (levels: apathy, empathy) and included random intercepts for participants and a correlated random slope for

empathic target within participant. Under the first hypothesis, intergroup empathic norms were expected to influence empathy for outgroup, but not ingroup members. This was formally tested by including empathic target as a fixed effect of interest in the model. If this expectation is supported, one should observe an intergroup empathic norm(s) x empathic target interaction revealing that neither of the norms influenced empathy for ingroup members while one or both influenced empathy for outgroup members.

In addition to running the model which included intergroup empathic norms as categorical variables, I examined participants' local adaptation to the most recent empathic feedback. Specifically, I regressed participants' empathic responses onto fixed effects for empathic target (levels: ingroup member, outgroup member), the most recent ingroup intergroup empathic feedback participants saw, and the most recent outgroup intergroup empathic feedback participants saw. This analysis also included random intercepts for participants and a correlated random slope for empathic target within participant.

In the first analysis, I found a significant main effect of empathic target ($F(1, 247) = 212.49$, $p < .001$), wherein participants empathised more with ingroup ($M = 2.76$, $SD = 1.83$) than outgroup members ($M = 1.11$, $SD = 2.42$), and a significant main effect of outgroup intergroup empathic norms ($F(1, 247) = 16.13$, $p < .001$), wherein participants exposed to outgroup members' intergroup empathy showed more empathy overall ($M = 2.27$, $SD = 2.18$) than participants exposed to outgroup members' intergroup apathy ($M = 1.60$, $SD = 2.36$). Importantly, these main effects were qualified by a significant empathic target x outgroup intergroup empathic norms interaction ($F(1, 247) = 14.81$, $p < .001$) which revealed that participants who witnessed outgroup members expressing empathy for their ingroup showed more empathy in response ($M = 1.70$, $SD = 2.33$) than participants who witnessed outgroup members expressing apathy for their ingroup ($M = 0.64$, $SD = 2.25$, $z = 4.75$, $p < .001$, see Figure 4.4). Empathy for ingroup members did not similarly vary across experimental

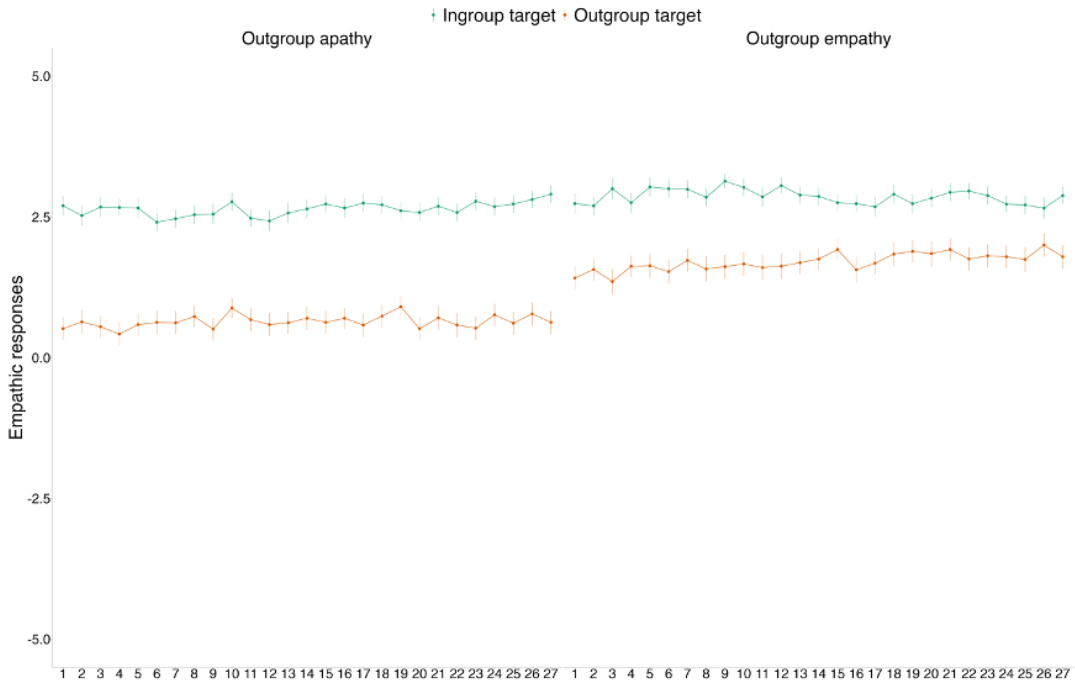
conditions where outgroup members showed intergroup empathy (M = 2.86, SD = 1.80) or intergroup apathy (M = 2.64, SD = 1.84, z = 1.41, p = .159). Furthermore, as is potentially already evident, participants' intergroup empathy bias was only reduced in the presence of outgroup members' intergroup empathy (b = 1.16, 95% CI = [0.76, 1.56], z = 7.57, p < .001) as opposed to outgroup members' intergroup apathy (b = 1.99, 95% CI = [1.59, 2.40], z = 13.05, p < .001) but was never fully eliminated (see Figure 4.4)[68]. Finally, the effect of ingroup intergroup empathic norms was trending in the same direction, but the two-way interaction with empathic target fell short of statistical significance (F(1, 247) = 3.21, p = .074)[69].

Figure 4.4

The influence of outgroup intergroup empathic norms on participants' intergroup empathy

[68] The intergroup empathy bias observed in the presence of outgroup members' intergroup empathy was not an artifact of analysing initial trials of the learning phase where participants did not know that outgroup members would be empathetic towards their ingroup. Specifically, this intergroup empathy bias persisted after excluding the first 20 trials (b = 1.08, 95% CI = [0.76, 1.39], z = 6.65, p < .001), 32 trials (b = 1.00, 95% CI = [0.68, 1.32], z = 6.08, p < .001) and even 42 trials of the learning phase (b = 0.97, 95% CI = [0.64, 1.30], z = 5.68, p < .001).

[69] This two-way interaction was significant in the model without the random slope for empathic target (F(1, 13299) = 46.08, p < .001). Under this model, participants empathised marginally more with outgroup members when ingroup members were also empathetic (M = 1.35, SD = 2.25) as opposed to apathetic towards the same outgroup (M = 0.99, SD = 2.43, z = 2.14, p = .064) while their intragroup empathy remained unaffected by these ingroup intergroup empathic norms (apathetic ingroup intergroup norm: M = 2.77, SD = 1.85; empathetic ingroup intergroup norm: M = 2.73, SD = 1.80, z = 0.26, p = .792).



Notes. Values on the y-axis represent the order of empathic targets (e.g., number 4 corresponds to the fourth ingroup and outgroup empathic target participants saw in the empathy learning task).

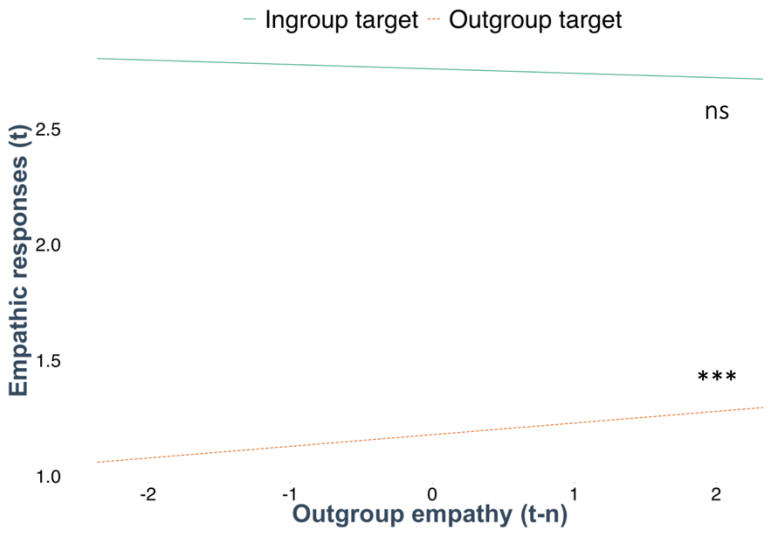
In the second analysis, I similarly observed a two-way interaction between the most recent outgroup intergroup empathic feedback and empathic target ($F(1, 12396) = 7.72$, $p < .001$). Specifically, the more empathy outgroup members showed for the most recent ingroup player in the past, the more participants empathised with the subsequent outgroup player ($b = 0.05$, 95% CI = [0.01, 0.09], $t = 2.81$, $p < .001$) while their empathy for the subsequent ingroup player remained unaffected by the outgroup intergroup empathic feedback that preceded it ($b = -0.02$, 95% CI = [-0.06, 0.02], $t = -1.09$, $p = .280$, see Figure 4.5). Unexpectedly, I also observed a two-way interaction between the most recent ingroup intergroup empathic feedback and the most recent outgroup intergroup empathic feedback ($F(1, 12424) = 8.41$, $p = .004$). In the absence of any significant simple slopes (all $ps > .110$), this interaction suggested a trend wherein the less empathy outgroup members displayed for the previous ingroup player, the more influence the most recent ingroup intergroup empathic feedback had on participants' empathic responses, regardless of who these empathic responses were directed towards (i.e., an ingroup or an outgroup empathic target). No other

significant effects were observed (all *ps* > .205)[70]. Somewhat unsurprisingly, both described effects disappeared after I excluded the initial 20 trials of the learning phase (all *ps* > .080). This suggests that the immediate increase in intergroup empathy in response to outgroup members' emotional reactions to ingroup members' outcomes mostly took place at the very beginning of the empathy learning task. Taken together, these results revealed that reciprocity enhanced empathy for outgroup members more than conformity.

[70] In contrast to the previous model, these results did not change regardless of how the random structure was specified.

Figure 4.5

The influence of the most recent outgroup intergroup empathic feedback on participants' intergroup empathy



Notes. The label *t* refers to the current trial on which participants are giving their empathic responses, while *n* refers to the number of trials that were in between the current trial and the most recent occurrence of outgroup intergroup empathic feedback. Values on the x-axis are standardised.

Public versus private displays of intergroup empathy. To examine how participants' displays of empathy changed across public and private settings, I regressed participants' empathic responses onto fixed effects for empathic target (levels: ingroup member, outgroup member), ingroup intergroup empathic norm (levels: apathy, empathy), outgroup intergroup empathic norm (levels: apathy, empathy) and response type (levels: private, public) and included random intercepts for participants and a correlated random slope for empathic target within participant. This model revealed two significant three-way interactions: (a) empathic target x ingroup intergroup empathic norm x response type interaction ($F(1, 17562) = 5.24$, $p = .022$), and (b) empathic target x outgroup intergroup empathic norm x response type ($F(1, 17562) = 20.10$, $p < .001$). Unpacking the first three-way interaction, regardless of which ingroup intergroup empathic norm participants were exposed to, their *intergroup* empathy went down once their responses became private

(apathetic ingroup intergroup empathic norm: b = -0.20, 95% CI = [-0.34, -0.07], z = 4.15, p < .001; empathetic ingroup intergroup empathic norm: b = -0.27, 95% CI = [-0.41, -0.14], z = 5.44, p < .001). On the other hand, regardless of which ingroup intergroup empathic norm participants were exposed to, their *intragroup* empathy remained the same once their responses became private (apathetic ingroup intergroup empathic norm: b = -0.06, 95% CI = [-0.19, 0.07], z = 1.16, p > .999; empathetic ingroup intergroup empathic norm: b = 0.10, 95% CI = [-0.03, 0.23], z = 2.01, p = .265). Unpacking the second three-way interaction, participants' private intergroup empathy was significantly reduced relative to their public intergroup empathy when they observed outgroup members' intergroup apathy (b = -0.41, 95% CI = [-0.54, -0.28], z = 8.21, p < .001), but not when they observed outgroup members' intergroup empathy (b = -0.07, 95% CI = [-0.20, 0.06], z = 1.40, p = .161). On the other hand, regardless of whether participants witnessed outgroup members' intergroup apathy or empathy, their own intragroup empathy remained unchanged across private and public settings (apathetic outgroup intergroup empathic norm: b = 0.08, 95% CI = [-0.06, 0.32], z = 1.53, p = .756; empathetic outgroup intergroup empathic norm: b = -0.03, 95% CI = [-0.16, 0.10], z = 0.65, p > .999).

I repeated the analysis outlined above after excluding the first 20 trials of the task's learning phase. It is important to recognise that when participants start the empathy learning task, they are not yet familiar with ingroup and outgroup intergroup empathic norms in their experimental condition, so their empathic responses on the initial trials of the task still greatly reflect their personal preferences and beliefs rather than norm adaptation. For this reason, I excluded these initial trials to assess how subsequent private displays of intergroup empathy compare to intergroup empathy participants display publicly once the operative intergroup empathic norms are fully known. The number of excluded trials was not chosen randomly, but instead represents the minimum number of trials at which all participants (N = 251) saw

at least 5 ingroup empathic targets and 5 outgroup empathic targets, allowing for sufficient learning of both intergroup empathic norms across all participants. Once these initial trials have been excluded, the same two three-way interactions were observed: (a) empathic target x ingroup intergroup empathic norm x response type interaction ($F(1, 13282.2) = 8.24$, $p = .004$), and (b) empathic target x outgroup intergroup empathic norm x response type interaction ($F(1, 13282.2) = 13.81$, $p < .001$). All results remained the same as in the first analysis with one important change. Specifically, in experimental conditions where outgroup intergroup empathy was normative, participants own intergroup empathy increased as the task progressed, and participants became aware of this norm. This increase meant that when participants' empathic responses subsequently became private, their intergroup empathy in fact reduced in comparison to the intergroup empathy they reached in the learning phase ($b = -0.12$, 95% CI = [-0.25, -0.01], $z = 2.40$, $p = .017$). This pattern was not an artifact of the specific number of trials I decided to exclude for this analysis but was even more pronounced after excluding 32 trials (half of the learning phase: $b = -0.17$, 95% CI = [-0.31, -0.04], $z = 3.34$, $p < .001$) and 42 trials (leaving 21 trials to match the length of the test phase: $b = -0.22$, 95% CI = [-0.36, -0.07], $z = 3.86$, $p < .001$). Two related details should be noted. First, the discrepancy between private and public intergroup empathy in conditions where outgroup members displayed intergroup empathy, although significant, was still smaller than the discrepancy between private and public intergroup empathy in conditions where outgroup members displayed intergroup apathy[71]. Second, although absolute intergroup empathy went down when reputational pressure was removed, participants' private intergroup empathy was still higher if they had previously seen outgroup members' intergroup empathy ($M = 1.63$, $SD = 2.45$) as opposed to intergroup apathy ($M = 0.23$, $SD = 2.56$, $z = 5.67$, $p < .001$). This

[71] This is indexed by a two-way outgroup intergroup empathic norm x response type interaction ($F(1, 6669.4) = 18.07$, $p < .001$) in the model which focused on participants' intergroup empathy.

suggests that the general influence of outgroup intergroup empathic norms remained equivalent across public and private intergroup empathy.

Finally, I zoomed in on the two incongruent conditions (i.e., conform-to-empathise and reciprocate-to-empathise) because in these conditions conformity and reciprocity competed for influence over participants' intergroup empathy. I hypothesised that participants in the conform-to-empathise condition would show a greater discrepancy between their public and private intergroup empathy in comparison to participants in the reciprocate-to-empathise condition because conformity to ingroup members' intergroup empathy is driven more by reputational concerns than reciprocity of outgroup members' intergroup empathy. To address this hypothesis, I regressed participants' empathic responses, either from the entire empathy learning task or after excluding the first 20 trials of the learning phase, onto fixed effects for empathic target (levels: ingroup member, outgroup member), experimental condition (levels: conform-to-empathise, reciprocate-to-empathise) and response type (levels: private, public) and included random intercepts for participants and a correlated random slope for empathic target within participant. Regardless of whether the first 20 trials were excluded, I observed a three-way interaction (full learning phase: $F(1, 8676) = 22.08$, $p < .001$; shortened learning phase: $F(1, 6572.6) = 20.31$, $p < .001$). When no trials were excluded, participants in the conform-to-empathise condition showed greater intragroup empathy ($b = 0.20$, 95% CI = [0.83, 0.01], $z = 2.74$, $p = .037$) and lower intergroup empathy ($b = -0.50$, 95% CI = [-0.69, -0.31], $z = 6.85$, $p < .001$) once their responses became private. On the other hand, participants in the reciprocate-to-empathise condition showed no differences between their private and public intragroup ($b = -0.07$, 95% CI = [-0.26, 0.12], $z = 0.96$, $p > .999$) and intergroup empathy ($b = -0.09$, 95% CI = [-0.28, 0.10], $z = 1.24$, $p = .216$). The results remained the same when the first 20 trials of the learning phase were excluded, with the exception that the discrepancy between private than public intragroup

empathy in the conform-to-empathise condition was no longer significant (b = 0.15, 95% CI = [-0.04, 0.35], z = 2.10, p = .214). In line with the analysis which examined all combinations of ingroup and outgroup intergroup empathic norms, in the reciprocate-to-empathise condition, the discrepancy between public and private intergroup empathy grew as more trials were excluded from the learning phase, thus allowing public intergroup empathy to unambiguously reflect adaptation to known intergroup empathic norms. However, this public-private discrepancy never reached statistical significance (20 excluded trials: b = -0.12, 95% CI = [-0.31, 0.07], z = 1.69, p = .184; 32 excluded trials: b = -0.16, 95% CI = [-0.36, 0.04], z = 2.09, p = .103; 42 excluded trials: b = -0.18, 95% CI = [-0.39, 0.04], z = 2.17, p = .090). Taken together, these findings suggest that participants tended to display less intergroup empathy when their responses were private as opposed to public[72], with this reduction being the lowest, although not completely absent across all analyses, when outgroup members displayed intergroup empathy. On the other hand, participants' intragroup empathy did not differ between private and public settings. Finally, conformity to ingroup members' intergroup empathy is more likely to wane in the absence of reputational concerns than reciprocity of outgroup members' intergroup empathy.

The moderating role of reputational concern in empathic norm compliance. Next, I examined whether participants' self-reported reputational concern predicted which intergroup empathic norm they followed in experimental conditions where conformity and reciprocity demanded different emotional responses. Focusing on responses in the learning phase of the task, I regressed participants' empathic responses onto fixed effects for empathic target (levels: ingroup member, outgroup member), experimental condition (levels: conform-to-empathise, reciprocate-to-empathise) and reputational concern and included random

[72] This was also evidenced by the two-way interaction between empathic target and response type in all reported analyses (all ps < .001), but for brevity, I only unpacked the highest order interactions in the main text.

intercepts for participants as well as a correlated random slope for empathic target within participant. Participants' reputational concern was standardised before its inclusion in the model. This model revealed a two-way empathic target x reputational concern interaction ($F(1, 120) = 6.17$, $p = .014$). On one hand, I observed no differences between participants with high and low reputational concern in how much they empathised with ingroup members ($b = -0.26$, 95% CI = [-1.00, 0.48], $z = -0.68$, $p = .495$). On the other hand, participants with low reputational concern empathised more with outgroup members than participants with high reputational concern ($b = 1.12$, 95% CI = [0.04, 2.19], $z = 2.03$, $p = .043$). This pattern of results does not support my initial hypothesis. Specifically, whereas I predicted that participants with high reputational concern would preferentially adhere to ingroup intergroup empathic norms over outgroup intergroup empathic norms, results revealed that these participants in fact showed lower intergroup empathy regardless of how ingroup members reacted to outgroup members' outcomes.

Accuracy of empathic norm perceptions. Finally, I examined whether participants' final perceptions of intragroup and intergroup empathic norms deviated from intragroup and intergroup empathic norms they were in fact exposed to[73]. Specifically, I recoded participants' learning check responses (see the section "Learning check questions") to a -5 (counter-empathy) to 5 (empathy) scale and compared these responses to the actual empathic norms (i.e., intergroup, and intragroup empathy ingroup and outgroup members showed on average) participants in each experimental condition were exposed to. To translate this to

[73] Please note that perceptions of *intergroup* empathic norms were of primary interest in Experiment 7, but I also examined participants' perceptions of *intragroup* empathic norms to assess their overall perceptions of ingroup and outgroup members' intergroup empathy bias (i.e., how much participants perceived ingroup and outgroup members to discriminate based on group membership in their empathy).

model-specific terms, I regressed perceived and actual empathic norms[74] onto fixed effects for class (levels: perceived empathic norms, actual empathic norms), empathic target (levels: ingroup member, outgroup member), empathiser (levels: ingroup members, outgroup members) and experimental condition (levels were the four conditions included in the experiment)[75], and included random intercepts for participants. This model yielded a significant four-way interaction ($F(3, 3737) = 4.85$, $p = .002$) and thus, I describe results from each experimental condition separately to facilitate interpretation (see Figure 4.6)[76].

First, participants assigned to the bidirectional intergroup empathy condition were accurate in their perceptions of ingroup members' intragroup empathy (actual vs. perceived ingroup intragroup empathy: $b = -0.16$, 95% CI = [-0.66, 0.33], $t = 1.02$, $p > .999$) and outgroup members' intragroup empathy (actual vs. perceived outgroup intragroup empathy: $b = -0.24$, 95% CI = [-0.73, 0.26], $t = 1.50$, $p > .999$). However, they tended to underestimate both ingroup members' intergroup empathy (actual vs. perceived ingroup intergroup empathy: $b = -0.82$, 95% CI = [-1.31, -0.32], $t = 5.17$ $p < .001$) and outgroup members'

[74] This simply means that the data were in the long format where each participant had both perceived and actual empathic norms for each empathiser x empathic target combination (i.e., 8 rows per participant). In other words, perceived and actual empathic norms were not analysed separately, but rather included in the same analysis.

[75] Given that I was interested in assessing participants' accuracy for each intra- and intergroup empathic norm within each experimental condition, experimental condition was included as a single fixed factor instead of having two fixed factors corresponding to ingroup intergroup empathic norms and outgroup intergroup empathic norms respectively.

[76] Of note, all patterns reported in this section (e.g., participants' tendency to report that ingroup and outgroup members showed intergroup bias in the bidirectional intergroup empathy condition despite there being none) were also observed when participants were asked to judge the feelings of average ingroup and outgroup members towards each minimal group using standard 100-point slider scales (e.g., "How [positively/negatively] do you think an average member of the [minimal ingroup/outgroup] feels towards the [minimal ingroup/outgroup]?"). This increases confidence that overestimation and underestimation of intergroup bias reported in this section are not artifacts of the emoji scale used to assess participants' perceptions of empathic norms (see "Learning check questions" for more details) but likely reflect biased representations of group relations participants formed based on individual ingroup and outgroup members' emotional reactions.

intergroup empathy (actual vs. perceived outgroup intergroup empathy: b = -1.08, 95% CI = [-1.57, -0.58], t = 6.82, p < .001) displayed in the task. In summary, these participants came to believe that both ingroup (M = 0.73, SD = 2.05) and outgroup members (M = 0.77, SD = 2.17) displayed intergroup empathy bias in the task even though in reality they displayed none (ingroup members' actual intergroup empathy bias: M = 0.08, SD = 0.40; actual vs. perceived ingroup intergroup empathy bias: t = 3.04, p = .007; outgroup members' actual intergroup empathy bias: M = -0.07, SD = 0.41; actual vs. perceived outgroup intergroup empathy bias: t = 3.90, p < .001).

Second, participants assigned to the conform-to-empathise condition were accurate in their perceptions of ingroup members' intragroup empathy (actual vs. perceived ingroup intragroup empathy: b = -0.24, 95% CI = [-0.77, 0.30], t = 1.39, p = .690) but they tended to underestimate outgroup members' intragroup empathy (actual vs. perceived outgroup intragroup empathy: b = -0.79, 95% CI = [-1.32, -0.25], t = -4.61, p < .001). Furthermore, these participants underestimated ingroup members' intergroup empathy (actual vs. perceived ingroup intergroup empathy: b = -1.64, 95% CI = [-2.17, -1.11], t = -9.61, p < .001) and overestimated outgroup members' intergroup empathy (actual vs. perceived outgroup intergroup empathy: b = 0.59, 95% CI = [0.06, 1.12], t = 3.45, p = .005). In summary, these participants overestimated ingroup members' levels of intergroup empathy bias (actual intergroup empathy bias: M = -0.02, SD = 0.36; perceived intergroup empathy bias: M = 1.39, SD = 1.39, t = 6.44, p < .001) but underestimated outgroup members' levels of intergroup empathy bias (actual intergroup empathy bias: M = 3.02, SD = 0.43; perceived intergroup empathy bias: M = 1.65, SD = 2.92, t = -6.31, p < .001).

Third, participants assigned to the reciprocate-to-empathise condition showed the exact opposite pattern from participants in the conform-to-empathise condition. Specifically, these participants were accurate in their perceptions of outgroup members' intragroup

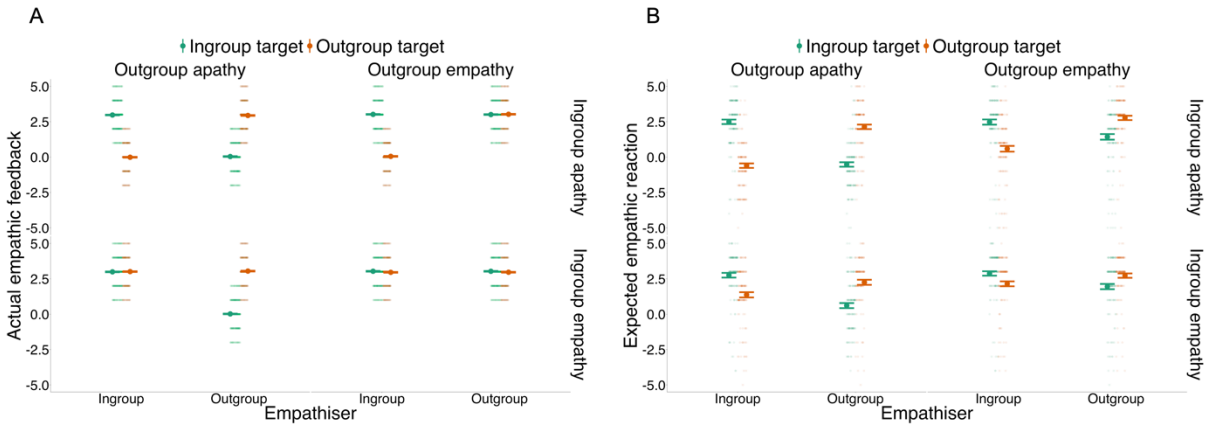
empathy (actual vs. perceived outgroup intragroup empathy: b = -0.25, 95% CI = [-0.78, 0.27], t = 1.52, p = .773) but they tended to underestimate ingroup members' intragroup empathy (actual vs. perceived ingroup intragroup empathy: b = -0.55, 95% CI = [-1.07, -0.02], t = 3.27, p = .012). Simultaneously, these participants overestimated ingroup members' intergroup empathy (actual vs. perceived ingroup intergroup empathy: b = 0.54, 95% CI = [0.02, 1.06], t = 3.23, p = .012) and underestimated outgroup members' intergroup empathy (actual vs. perceived outgroup intergroup empathy: b = -1.58, 95% CI = [-2.10, -1.05], t = 9.41, p < .001). In summary, participants in the reciprocate-to-empathise condition underestimated ingroup members' levels of intergroup empathy bias (actual intergroup empathy bias: M = 2.98, SD = 0.37; perceived intergroup empathy bias: M = 1.89, SD = 2.99, t = -5.08, p < .001) but overestimated outgroup members' levels of intergroup empathy bias (actual intergroup empathy bias: M = 0.02, SD = 0.38; perceived intergroup empathy bias: M = 1.34, SD = 2.50, t = 6.18, p < .001).

Fourth, participants assigned to the bidirectional intergroup apathy condition showed a tendency to underestimate empathy shown by both sets of empathisers towards both sets of empathic targets (actual vs. perceived ingroup intragroup empathy: b = -0.48, 95% CI = [-1.00, -0.04], z = -2.91, p = .022; actual vs. perceived outgroup intragroup empathy: b = -0.81, 95% CI = [-1.33, -0.29], z = -4.91, p < .001; actual vs. perceived ingroup intergroup empathy: b = -0.60, 95% CI = [-1.12, -0.08], z = -3.64, p = .003; actual vs. perceived outgroup intergroup empathy: b = -0.56, 95% CI = [-1.07, -0.04], z = -3.39, p = .006). Given that these participants underestimated empathy in all combinations of empathisers and empathic targets to a similar extent, their perceptions of ingroup (M = 3.10, SD = 2.29) and outgroup members' (M = 2.66, SD = 2.67) intergroup empathy bias were accurate overall (ingroup members' actual intergroup empathy bias: M = 2.98, SD = 0.37; actual vs. perceived ingroup intergroup empathy bias: t = 0.57, p > .999; outgroup members' actual intergroup empathy

bias: M = 2.91, SD = 0.39; actual vs. perceived outgroup intergroup empathy bias: t = -1.19, p = .937).

Figure 4.6

Actual versus perceived intra- and intergroup empathic norms



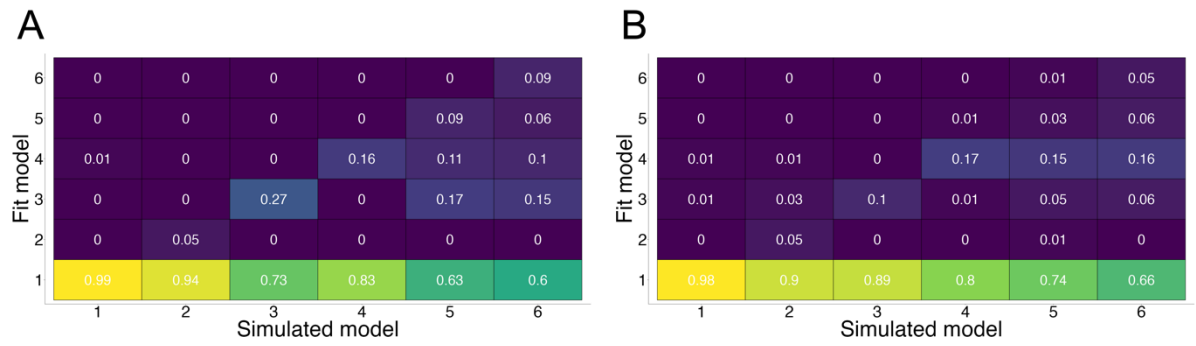
Notes. (A) Actual intra- and intergroup empathic norms in each experimental condition. Lighter dots in the background represent individual ingroup and outgroup members' emotional reactions throughout the empathy learning task (see Figure 4.3C for possible values), while solid dots and error bars in the foreground represent means and standard errors. (B) Perceptions of intra- and intergroup empathic norms participants in each experimental condition reported after completing the empathy learning task. Lighter dots in the background represent participants' individual responses, while solid dots and error bars in the foreground represent means and standard errors.

Biases in observational learning. The results from accuracy analyses showed that observational learning consistently led to erroneous perceptions of both intra- and intergroup empathic norms across all experimental conditions, which justified further investigation of underlying learning biases. Therefore, I fitted six reinforcement learning (RL) models (Lockwood & Klein-Flügge, 2021) which implemented different hypotheses about the mechanisms underlying the formation of erroneous perceptions of intergroup empathic norms observed in the experiment. However, as can be seen in Figure 4.7, these models were not distinguishable even in the best-case scenario when the data are simulated on the basis of a known model. Specifically, the simplest model (i.e., Model 1) consistently had the highest probability of being the best fitting model regardless of the true model based on which data

were simulated, regardless of whether the original parameter estimation procedure, which determined the range of parameters used in data simulation, adopted priors to regularise model parameters and regardless of the exact model comparison method (AIC or BIC). Given that Experiment 7 sought to determine which set of learning biases, represented in different learning models, best describes the responses in the empathy learning task as well as the final perceptions of intergroup empathic norms, the inability to recover known models renders any further interpretation of model comparison results misleading. Instead, these results call for rethinking of the fitted models as well as the suitability of the empathy learning task to generate data rich enough for addressing this type of learning questions. In the discussion of Experiment 7, I elaborate on these points and suggest the next steps to address this issue. The full description of the six fitted models as well as the model recovery steps can be found in Appendices 4.1 and 4.2 respectively.

Figure 4.7.

Confusion matrices quantifying the probabilities that each model is the best fit to the data generated from each model



Notes. These confusion matrices represent the probabilities that each model is the best fit to the data generated from each model (i.e., *p* (fit model | simulated model)). For instance, the number 0.6 in the lower right cell of sub-figure A represents the probability that Model 1 (i.e., the simplest model) is the best fit to the data generated from Model 6 (i.e., the most complex model). (A) Confusion matrix obtained when the data were simulated using parameters that matched the range of values derived from *non-regularised* parameter estimation (see Appendices 4.1 and 4.2 for details). (B) Confusion matrix obtained when the data were simulated using parameters that matched the range of values derived from *regularised* parameter estimation (see Appendices 4.1

and 4.2 for details). In both sub-figures, AIC was used for model comparison, but the results remained the same when the BIC was adopted.

Secondary analyses. Alongside addressing hypotheses listed in the introduction of Experiment 7, I also conducted a series of secondary analyses. First, in addition to examining the moderating role of self-reported reputational concern, I examined whether participants' perceptions of authenticity and/or appropriateness of intergroup empathic norms could discriminate between those who chose to align with ingroup intergroup empathic norms and those who chose to align with outgroup intergroup empathic norms in incongruent experimental conditions. These results are reported in the section "Exploratory moderation analyses" below. Second, I investigated how exposure to different intergroup empathic norms influenced participants' own feelings towards and identification with the minimal ingroup and the minimal outgroup. These results are reported in the section "The influence of intergroup empathic norms on one's feelings and identification" below.

***Exploratory moderation analyses*.** To examine whether individual differences in preferences for conformity versus reciprocity can be predicted by individual differences in how (a) authentic and/or (b) appropriate ingroup and outgroup intergroup empathic norms were perceived to be, I fitted two linear mixed models. The first model included experimental condition (levels: conform-to-empathise, reciprocate-to-empathise) and difference in intergroup empathic norms' authenticity (i.e., how authentic ingroup intergroup empathic norm was perceived versus how authentic outgroup intergroup empathic norm was perceived) as fixed factors, and the second model included experimental condition (levels: conform-to-empathise, reciprocate-to-empathise) and difference in intergroup empathic norms' appropriateness (i.e., how appropriate ingroup intergroup empathic norm was perceived versus how appropriate outgroup intergroup empathic norm was perceived) as fixed factors. Both models included random intercepts for participants. Using information criteria, these

exploratory models were compared to the baseline model which included only experimental condition as a fixed effect.

The inclusion of perceived intergroup empathic norms' appropriateness improved model fit in comparison to the baseline model (ΔAIC = -13, $p < .001$), while no such improvement was observed upon inclusion of perceived intergroup empathic norms' authenticity (ΔAIC = 0, $p = .118$). Zooming in on the model with perceived intergroup empathic norms' appropriateness, I observed that the more participants judged a certain intergroup empathic norm as appropriate, the more they tended to adhere to it. First, in the conform-to-empathise condition, participants who thought that ingroup members' intergroup empathy was more appropriate than outgroup members' intergroup apathy tended to show greater intergroup empathy themselves, therefore conforming more, than participants who thought the opposite ($b = 0.64$, 95% CI = [0.17, 1.11], $t = 2.68$, $p = .007$). Second, in the reciprocate-to-empathise condition, participants who thought that ingroup members' intergroup apathy was more appropriate than outgroup members' intergroup empathy tended to show lower intergroup empathy themselves, therefore conforming more, than participants who thought the opposite ($b = -0.73$, 95% CI = [-1.18, -0.28], $z = -3.23$, $p = .001$). However, given that judgments of intergroup empathic norms' appropriateness were provided at the end of the experiment, one cannot rule out the possibility that participants used these judgments as means to justify their own choices in the empathy learning task. Specifically, while one interpretation could be that participants adhered to a certain intergroup empathic norm because they thought this norm was more appropriate, it is equally possible that participants subsequently judged a certain intergroup empathic norm as more appropriate because they had previously adhered to it. Thus, the direction of this interpretation remains ambiguous and will not be further discussed. Nonetheless, this finding invites future research to explicitly examine the interaction between "personal" intergroup empathic norms (i.e., how much an

individual considers empathy an appropriate and desirable emotional response in each given intergroup context) and "social" intergroup empathic norms (i.e., how much relevant others consider empathy as appropriate and desirable emotional response in each given intergroup context, see Sherman & Cohen, 2006; Zaki, 2014).

The influence of intergroup empathic norms on one's feelings and identification. To examine how intergroup empathic norms influenced participants' feelings towards and identification with minimal ingroup and outgroup members, I fitted a linear mixed model with respective dependent variables, three fixed effects – target (levels: minimal ingroup, minimal outgroup), ingroup intergroup empathic norm (levels: apathy, empathy), and outgroup intergroup empathic norm (levels: apathy, empathy) – and random intercepts for participants. Both models yielded similar results. First, across different intergroup empathic norms, participants reported more positive feelings towards and greater identification with minimal ingroup members (feelings: $M = 79.13$, $SD = 18.56$; identification: $M = 11.38$, $SD = 2.40$) than minimal outgroup members (feelings: $M = 58.18$, $SD = 22.91$, $t = 12.04$, $p < .001$; identification: $M = 9.06$, $SD = 2.75$, $t = 11.49$, $p < .001$). Second, participants who were exposed to outgroup members' intergroup empathy reported more positive feelings for and greater identification with the minimal outgroup (feelings: $M = 62.94$, $SD = 23.09$; identification: $M = 9.63$, $SD = 2.81$) than participants' who were exposed to outgroup members' intergroup apathy (feelings: $M = 53.46$, $SD = 21.81$, $t = 3.66$, $p < .001$; identification: $M = 8.49$, $SD = 2.57$, $t = 3.58$, $p < .001$). Finally, participants in experimental conditions where intergroup apathy was normative among outgroup members did not only report more negative feelings towards the minimal outgroup but also more positive feelings towards the minimal ingroup ($M = 81.71$, $SD = 16.73$) in comparison to participants in experimental conditions where intergroup empathy was normative among outgroup members ($M = 76.53$, $SD = 19.97$, $t = 1.96$, $p = .051$). This pattern was not replicated in participants'

identification responses where exposure to outgroup members' intergroup empathy or intergroup apathy made no difference to how much participants identified with their minimal ingroup (apathetic outgroup intergroup norm: $M = 11.53$, $SD = 2.36$; empathetic outgroup intergroup norm: $M = 11.23$, $SD = 2.45$, $t = 0.91$, $p = .362$).

Discussion

Although people often feel reduced empathy for outgroup members, an emerging consensus is that portraying intergroup empathy as normative among ingroup or outgroup members can help increase it. For instance, people have been shown to empathise more with outgroup members when their ingroup members are also empathetic (*conformity*, Saguy et al., 2015; Tarrant et al., 2009) or when outgroup members are empathetic towards one's ingroup (*reciprocity*, Gubler et al., 2015; Hein et al., 2016; Nadler & Liviatan, 2006). However, past work has only looked at ingroup and outgroup intergroup empathic norms in isolation from each other, making it difficult to predict how displays of intergroup empathy vary across social environments where cues of both intergroup empathic norms are available simultaneously and promote either congruent or incongruent emotional responses. To address this question, in Experiment 7, I had participants play an online game in which they had to learn about both ingroup and outgroup intergroup empathic norms by observing individual ingroup and outgroup members' emotional reactions to each other's outcomes. In this game, conformity and reciprocity were either aligned (e.g., both promoted apathy or empathy) or misaligned (e.g., one promoted apathy, the other one empathy). The main hypothesis was that people would both conform to ingroup intergroup empathic norms and reciprocate outgroup intergroup empathic norms, with reciprocity more strongly affecting intergroup empathy than conformity.

Experiment 7 provided evidence that reciprocity outperforms conformity in promoting intergroup empathy. Specifically, in the context of minimal groups, participants

primarily adjusted their intergroup empathy to match outgroup members' emotional reactions towards their ingroup. If outgroup members displayed empathy as opposed to apathy in response to ingroup members' outcomes, participants tended to show more empathy across group lines, with this increase in intergroup empathy corresponding to 1.06 points on an 11-point scale[77]. While participants' reciprocity of outgroup intergroup empathic norms was consistently documented, participants' conformity to ingroup intergroup empathic norms was only indexed by statistical trends and fell short of statistical significance. Therefore, although ingroup intergroup empathic norms might influence intergroup empathy in isolation, their influence might wane once they are put into opposition with outgroup intergroup empathic norms. Furthermore, witnessing outgroup members displaying intergroup empathy as opposed to apathy resulted in 17.7% improvement in feelings towards and 13.4% increase in identification with the outgroup, while witnessing ingroup members displaying intergroup empathy as opposed to apathy made no difference to how participants felt towards and how much they identified with the outgroup. In summary, Experiment 7 provides the first evidence that outgroup intergroup empathic norms might be a more important driver of people's motivation to display empathy for outgroup members than ingroup intergroup empathic norms, with their influence potentially transcending displays of intergroup empathy to generate more positive feelings towards outgroup members. Nonetheless, these results do not allow for immediate generalisation to real-world intergroup contexts or more realistic empathy-demanding situations but should instead be carefully interpreted in the context of

[77] Since Experiment 7 was the first one to contrast conformity and reciprocity in the domain of intergroup empathy and the first one to assess the influence of intergroup empathic norms using a subtle learning paradigm, no smallest (unstandardised) effect size of interest was specified in advance. Thus, while I have strived to transparently report unstandardised effect sizes (i.e., understandardised mean differences) throughout this chapter to make it easier for the reader to assess the results, I refrain from subjective post hoc evaluations of these results' practical meaningfulness.

the current experiment. On this note, a few features of Experiment 7 are important to mention.

First and foremost, Experiment 7 relied on minimal groups, which lack the conceptual richness and the recognised meaningfulness of real-world groups, making it reasonable to assume that minimal ingroups likely exerted less influence on people's displays of intergroup empathy than real-world ingroups would have done. Although conformity is likely to increase in the presence of real-world ingroups that people genuinely identify with, this does not necessarily mean that the balance would be tilted in favour of conformity over reciprocity, especially in light of abundant evidence showing that people are eager to reciprocate the treatment they believe their real-world ingroups receive from real-world outgroups (e.g., Kteily et al., 2016; Landry et al., 2022; Mernyk et al., 2022). More work is thus needed to clarify to what extent the current results generalise to various real-world environments and to identify relevant moderators determining which intergroup empathic norms are preferentially followed. Additionally, I invite future researchers to consider combining the study of minimal groups with the study of existing groups to investigate whether those who prefer adhering to a certain norm in a minimal intergroup environment maintain the same preference among real-world groups. Any continuity between minimal-group and real-group effects would underscore the utility of reconceptualising the minimal group paradigm as an individual differences-paradigm used to index people's general "groupiness" (see Chapter 2 for more details on this idea).

Second, another feature of Experiment 7 which potentially amplified the influence of reciprocity over conformity was that participants did not suffer any consequences (i.e., punishment) regardless of how much they deviated from ingroup intergroup empathic norms in the empathy learning task. Specifically, minimal ingroup members always responded empathically to participants' outcomes, regardless of whether participants behaved in line

with or diverged from the rest of the group. However, it is known that people commonly punish others who engage in norm-violating behaviour (Crockett et al., 2014; Molho et al., 2020; van Dijk et al., 2015), even when this norm-violating behaviour is prosocial (e.g., displaying higher generosity than the rest of the group, see Kawamura & Kusumi, 2020; Parks & Stone, 2010). In light of these findings, one could modify the existing empathy learning task to allow ingroup members to punish participants when they violate ingroup intergroup empathic norms (e.g., when they display intergroup empathy despite knowing that the rest of the group does not or when they refuse to display intergroup empathy despite having witnessed the rest of the group do it). The advantage of the empathy learning task is that it is highly customisable and thus able to incorporate different forms of direct or indirect punishment (e.g., direct confrontation, gossip, social exclusion, Molho et al., 2020). To illustrate, if a researcher is interested in the effects of direct confrontation, they can simulate a situation where one or multiple ingroup members scold the participant for their norm violation or have ingroup members change their emotional responses to participants' outcomes from empathy to apathy once a violation of ingroup intergroup empathic norms has been detected. In summary, introducing ingroup-bounded punishment is likely to motivate participants to prefer conformity over reciprocity, at least in public.

Regarding the potential discrepancy between private and public expressions of intergroup empathy, future research should determine whether persistence of conformity in private depends on how ingroup intergroup empathic norms are enforced (e.g., whether participants' conformity is reduced in private only in the absence of punishment while punishment maintains the same level of conformity across private and public settings). Investigating the role of punishment is particularly important given the prevalence of intergroup contexts (e.g., contexts characterised by present or past competition and conflict) where displays of intergroup empathy can be counter-normative to such an extent that

publicly expressing empathy for the outgroup might trigger various forms of punishment aimed to restore empathisers' norm compliance and prevent interference of empathy with competitive goals. While it is important to understand how different forms of punishment reinforce conformity, future research should also aim to identify conditions under which refusal to conform to norms of intergroup apathy (or any other outgroup-oriented norms) could result in norm change rather than ostracism.

Third and finally, the emoji scale (see Figure 4.3) adopted in Experiment 7, however illustrative, is far from how empathy is usually experienced and displayed in real life. One important difference between "real life empathy" and "emoji empathy" is that "real life empathy" is commonly perceived as aversive and cognitively effortful, leading people to avoid it (Cameron et al., 2019). This raises the possibility that had intergroup empathy required more effort (e.g., vicarious resonance and/or lengthy expression of support), participants would have engaged less in it than they did in the current set-up of Experiment 7, especially when norms of intergroup apathy were available, thus providing a justification for intergroup empathy avoidance. To address this possibility, future research should expand the assessment of the influence of intergroup empathic norms beyond undemanding self-report measures of intergroup empathy to include more effortful indices of empathy and prosocial orientation (e.g., support provision). Adopting a different perspective, considering the effortlessness of expressing intergroup empathy on the emoji scale, it is somewhat surprising that participants continued to display intergroup empathy bias even in the experimental condition where both ingroup and outgroup members showed no such bias (i.e., bidirectional intergroup empathy condition). This persistence of intergroup empathy bias in the bidirectional intergroup empathy condition might be linked to biased learning of intergroup empathic norms, as indexed by participants' erroneous assumptions that ingroup and outgroup members expressed intergroup empathy bias when they in fact did not. Thus, if

participants in the bidirectional intergroup empathy condition wrongfully perceived others to be displaying intergroup empathy bias, that could explain why they kept displaying one themselves. Nonetheless, it should not be understated that participants in Experiment 7 expected to interact with and depend on ingroup members in the later section of the experiment which could have acted as an incentive to display ingroup preference in the empathy learning task. This is not to say that participants were acting strategically, but rather that Experiment 7 imbued minimal groups with coalitional meaning, which has previously been shown to underpin different ingroup preferences (Misch et al., 2021).

The role of reputational concerns. Next, I examined how participants' displays of intra- and intergroup empathy changed when reputational concerns associated with the public nature of empathic responses were eliminated. Two hypotheses were initially offered in relation to this research question. Motivated by the set-up of the empathy learning task, the first hypothesis was that public and private expressions of both intra- and intergroup empathy would not differ because displays of empathy would start to act more as a habit than a deliberate, goal-directed emotional response (see Mazar & Wood, 2018; Mazar & Wood, 2022; Wood & Neal, 2007; Wood & Rünger, 2016). In other words, although people may initially display intra- and intergroup empathy to achieve a goal (e.g., to be accepted by others, to be nice to others), after repeating the same displays over and over again, people might start to associate contextual cues (i.e., outgroup members) with responses (i.e., particular apathetic/empathetic reactions), such that they later display intra- and intergroup empathy automatically (Hackel et al., 2019). The second hypothesis did not centre on habit formation but instead on reputational concerns that were expected to arise in experimental conditions where conformity and reciprocity were competing for influence over participants' intergroup empathy. Specifically, the second hypothesis predicted that conformity to ingroup members' unidirectional intergroup empathy is more likely to be driven by reputational

concerns than reciprocity of outgroup members' unidirectional intergroup empathy. Support for the second hypothesis would have been provided by participants assigned to the conform-to-empathise condition (i.e., ingroup empathy for the outgroup, outgroup apathy for the ingroup) showing a greater discrepancy between private and public intergroup empathy than participants assigned to the reciprocate-to-empathise condition (i.e., ingroup apathy for the outgroup, outgroup empathy for the ingroup).

In terms of results, Experiment 7 supported the intragroup, but not the intergroup part of the first hypothesis. Specifically, participants expressed the same level of intragroup empathy across private and public settings, but less intergroup empathy in private than in public settings, with the reduction in average intergroup empathy being subtle and ranging from 0.20 to 0.41 on an 11-point scale. The discrepancy between private and public expressions of intergroup empathy was reduced, but not entirely eliminated, in experimental conditions where outgroup members displayed intergroup empathy as opposed to intergroup apathy. Importantly, despite displays of intergroup empathy decreasing in absolute terms when responses became private, it should not be lost in translation that outgroup intergroup empathic norms similarly influenced both private and public expressions of intergroup empathy. In other words, regardless of whether responses were private or public, participants displayed more empathy for outgroup members if those outgroup members had previously expressed empathy as opposed to apathy for participants' ingroup. In summary, Experiment 7 documented that removing reputational concerns stemming from the visibility of one's emotional reactions is likely to reduce displays of intergroup, but not intragroup, empathy across normative contexts.

A few additional points elaborate on this finding. First, it might seem unintuitive that private intergroup empathy was reduced even when intergroup empathy was counter-normative in public. Why would people enhance public displays of intergroup empathy in

social contexts in which others do not look favourably on these displays to begin with? The most likely explanation is that the general perception of empathy as desirable and moral (Fowler et al., 2021; Schumann et al., 2014) continued to motivate people to upregulate their displays of empathy in the presence of others. However, this upregulation was subtle, with participants' public displays of intergroup empathy remaining within the range of the acceptable emotional reactions presented in their experimental condition. To understand this point, one should remember that intergroup empathic norms were illustrated not with a single, consistent emotional reaction (i.e., a single emoji) but rather a range of emotional reactions (i.e., a range of five emojis displayed with different probabilities, see Figure 4.3). This meant that participants assigned to experimental conditions which involved intergroup apathy could see others' emotional reactions ranging from slight counter-empathy to slight empathy. In this context, empathy's general desirability might have nudged participants to err towards showing slightly higher intergroup empathy in public while still remaining within the range of normative emotional reactions. In this way, participants avoided violating a norm while still acquiring indirect reputational benefits from their low-intensity empathic displays. Second, regarding participants' displays of intragroup empathy, one possibility is that participants' identification with their minimal ingroup increased after witnessing repeated displays of intragroup empathy, and that this increased identification, instead of reputational concerns, underpinned the persistence of intragroup empathy across private and public settings.

The habitual nature of intragroup empathic displays provides another possible account of their persistence across private and public settings. On this note, although Experiment 7 was not designed to delineate the role of habits and goal pursuit in displays of intra- and intergroup empathy, its set-up bears some interesting resemblance to outcome devaluation procedures that are commonly adopted to achieve this delineation (Adams & Dickinson,

1981; Coutureau & Killcross, 2003; Gillan et al., 2015; Robbins & Costa, 2017; Yin et al., 2004). A typical outcome devaluation procedure includes two steps. In the first step, people (or animals) learn to perform an action to secure an outcome and in the second step, which takes place only after the action is fully established, the value of this outcome is reduced. Thus, if a person continues to perform the learnt action despite the reduced value of the outcome, their action is considered to be *habitual*, and if the person reduces their performance of the learnt action in response to the reduced value of the outcome, their action is considered to be *goal-directed* (Gillan et al., 2015). With its removal of reputational benefits in the test phase, the empathy learning task broadly parallels the outcome devaluation procedure, and thus, raises a question of whether displays of intragroup empathy might reflect habitual as opposed to goal-driven behaviour. Nonetheless, on its own, the present empathy learning task is only able to pose this question but not to offer responses as it does not satisfy other criteria relevant for the study of habits (e.g., it does not involve explicit instrumental learning from rewards, nor does it devalue other possible reinforcers such as anticipation of interdependence). Future research should thus systematically examine whether ingroup-oriented and outgroup-oriented behaviours, including displays of empathy, vary in the extent to which they can be made to operate like habits (see also Amodio & Berg, 2022; Devine, 1989).

Moving onto the second hypothesis, Experiment 7 provided evidence in its favour. As expected, conformity to ingroup members' intergroup empathy was driven by reputational concerns more than reciprocity of outgroup members' intergroup empathy. This was indicated by participants in the conform-to-empathise condition showing a significant reduction in their private as opposed to public expressions of intergroup empathy, with this reduction corresponding to half a point on an 11-point scale, while the difference between private and public expressions of empathy was not significant in the reciprocate-to-empathise

condition. However, before elaborating on this finding, it is important to note that a similar conclusion was not supported using a different approach to examining the role of reputational concerns.

Under this second approach, I asked participants to report how much reputational concern they felt in relation to their minimal ingroup members. I then used these responses to predict which source of normative influence participants would decide to follow when conformity and reciprocity are misaligned. If reputational concerns motivate conformity but not reciprocity, in the conform-to-empathise condition, participants who report high reputational concern should show more intergroup empathy than participants who report low reputational concern, while in the reciprocate-to-empathise condition, participants who report high reputational concern should show less intergroup empathy than participants who report low reputational concern (see Romano & Balliet, 2017 for a similar result). In other words, high reputational concern should be associated with a greater tendency to conform, regardless of whether conformity involves displaying more or less intergroup empathy[78]. This hypothesis was not backed up by Experiment 7. Specifically, regardless of whether intergroup empathy or intergroup apathy was normative among ingroup members, participants with low reputational concern (i.e., participants who reported being less concerned about how they were evaluated by ingroup members) showed higher intergroup empathy. I speculate that multiple factors could have contributed to this result not aligning with the previous result contrasting private and public displays of intergroup empathy. First is the possibility that the items in the measure of reputational concern were interpreted differently depending on the participants' experimental condition and the degree of

[78] To align with the previous result contrasting private and public displays of intergroup empathy, high reputational concern should at least be associated with a greater tendency to conform to ingroup members' intergroup empathy in the conform-to-empathise condition.

conformity they themselves displayed in the empathy learning task. To illustrate, let me focus on the conform-to-reciprocate condition and use a sample item: "During the previous card game, I thought about how my group members would think about me". One can imagine both participants who conformed and who did not conform to ingroup members' intergroup empathy responding to this item with a high value, but for different reasons. On one side, a participant who conformed could have responded to this item with a high value to reflect that they were thinking about ingroup members' opinion of them and altering their displays of intergroup empathy to ensure this opinion was positive (i.e., "I care about how my group members think of me, thus I am going to behave how they behave."). On the other side, a participant who did not conform could have similarly responded to this item with a high value to reflect that they were thinking about ingroup members' opinion of them after they had decided not to side with them in their displays of intergroup empathy (i.e., "I am not behaving in line with the rest of the group, what might they be thinking about me now?"). It should be noted, however, that Experiment 7 used the same measure of reputational concern as Romano and Balliet (2017), whose work the starting hypothesis was based on, rendering the above speculation unable to account for why the current results differ from those previous authors' (2017) results. Second, it is important to note two differences between Romano and Balliet's (2017) experiments and Experiment 7. The first difference is that participants in Experiment 7 responded to both ingroup and outgroup members, while participants in Romano and Balliet's (2017) experiments did not respond to their ingroup members, but only to a single outgroup member. This means that while participants in Experiment 7 could have secured ingroup members' positive evaluation by showing empathy for ingroup members' outcomes as opposed to conforming to ingroup intergroup empathic norms, participants in the Romano and Balliet's (2017) experiments did not have this opportunity, potentially leading to an increase in the influence of reputational concerns on conformity to ingroup *intergroup*

norms. The second difference is that Romano and Balliet (2017) adopted an extensive procedure to ensure that participants developed a sense of group identification and cohesion prior to assessing conformity, while Experiment 7 had participants complete the empathy learning task almost immediately after group allocation, which could have led to reduced reputational concerns. In summary, despite the inconsistencies between the two approaches, Experiment 7 does provide at least some evidence that reputational concerns are more likely to motivate people to conform to ingroup members' intergroup empathy than to reciprocate outgroup members' intergroup empathy, but I urge future research to continue examining the role of this motive.

Learning of empathic norms. Thus far, I have discussed how intergroup empathic norms can encourage or discourage displays of intergroup empathy and how these displays might change across public settings, where one's reputation is at stake, and private settings, where one's actions are shielded from others' evaluation. Another important finding is that people form perceptions of intergroup empathic norms in a biased manner, which results in these perceptions not being an accurate reflection of true intergroup empathic norms people were in fact exposed to. While Experiment 7 could not provide a computational account of learning mechanisms underlying inaccurate perceptions of empathic norms due to model recovery issues, assessing participants' final perceptions of intergroup empathic norms and the patterns of inaccuracy in these perceptions might provide initial model-independent indications that certain biases in observational learning might have been engaged.

First, participants in the two congruent experimental conditions – bidirectional intergroup empathy and bidirectional intergroup apathy condition – formed perceptions of intergroup empathic norms which underestimated how much intergroup empathy ingroup and outgroup members displayed in the empathy learning task. This underestimation of intergroup empathy ranged from 0.48 to 1.08 points on an 11-point scale and might have

stemmed from the combination of two learning biases: (a) participants' negative priors about intergroup empathy, and (b) participants' negativity bias wherein they updated perceptions of intergroup empathic norms more on the basis of negative social information (i.e., observing a group member displaying less intergroup empathy than one expected) than positive social information (i.e., observing a group member displaying more intergroup empathy than one expected). This possibility is supported by data simulation. Specifically, data simulated on the basis of a computational model which incorporated these two biases (i.e., Model 3 in Appendix 4.1) was able to reproduce the pattern of underestimation observed in real participants' perceptions of intergroup empathic norms (see Appendix 4.3).

Second, participants in the two incongruent experimental conditions – conform-to-empathise and reciprocate-to-empathise – showed a tendency to overestimate the intergroup empathy shown by the unempathetic group, with this overestimation corresponding to 0.54-0.59 points on an 11-point scale, and to underestimate the intergroup empathy shown by the empathetic group, with this underestimation corresponding to 1.58-1.64 points on an 11-point scale. This pattern of results is expected to arise if participants display a reciprocity bias, wherein they generalise one group's emotional reactions to the other group. For instance, in the conform-to-empathise condition, witnessing outgroup members expressing intergroup apathy might have generalised to ingroup members due to expectations of reciprocity and thus interfered with accurate learning of ingroup members' emotional reactions, resulting in underestimation of their intergroup empathy. Simultaneously, witnessing ingroup members expressing intergroup empathy might have similarly generalised and interfered with learning of outgroup members' emotional reactions, thus leading to overestimation of their intergroup empathy. Data simulation on the basis of a computational model which assumed reciprocity bias alongside negativity bias (i.e., Model 5 in Appendix 4.1) was able to reproduce the combined pattern of underestimation and overestimation observed in real participants'

perceptions of intergroup empathic norms (see Appendix 4.3). Finally, it is important to note that accuracy results also suggested that learning of intra- and intergroup empathic norms might have similarly interfered (e.g., participants underestimated outgroup members' intragroup empathy when they witnessed them display intergroup apathy). However, models which combine the learning of intra- and intergroup empathic norms are likely too complex for the currently available data and thus, present an avenue for future research and are not further discussed.

Taken together, these results indicate that negatively biased and inaccurate perceptions of intergroup empathic norms cannot be entirely attributed to disproportionate exposure to negative intergroup content because people are likely to form these perceptions even after being exclusively exposed to impartial intergroup interactions. However, considering the model recovery issues, the discussed learning biases (i.e., negative priors, negativity bias, reciprocity bias) and their proposed association with the observed patterns of inaccuracy should only be used as a starting point to develop hypotheses concerning how perceptions of intergroup empathic norms are formed. On the note of model recovery, there are multiple factors that might have jointly contributed to the inability to recover simulated models. First, the empathy learning task, specifically designed and programmed for Experiment 7, stepped away from the standard binary- or multi-choice tasks commonly used in observational and reinforcement learning literature (see Cutler et al., 2021; Hackel et al., 2022; Selbing et al., 2014 for examples), making it difficult to predict whether the task would generate data that was rich enough to allow identification of hypothesised models and parameters (Wilson & Collins, 2019). Second, although the computational models fitted in Experiment 7 might have been psychologically intuitive and might have generated patterns of underestimation and overestimation observed in real participants' perceptions of intergroup empathic norms, it might still be the case that reinforcement learning (RL) models are

suboptimal for the data that were analysed. Specifically, RL models aim to describe how the value of different actions is learnt through making choices and experiencing rewards (Sutton & Barto, 1998), which does not fully correspond to how learning of intergroup empathic norms was structured in the empathy learning task. That is, while RL models are primarily used to describe the *active* process of action-outcome learning (Hackel et al., 2019), learning in Experiment 7 was rather *passive* and relied on simple exposure to the behaviours of others without these behaviours being a direct outcome of participants' actions (i.e., their emoji choices). A third plausible contributor to the failure of model recovery is that simulated parameters in Models 2-6 (see Appendix 4.1) were too similar to each other, resulting in the simplest model (i.e., Model 1 in Appendix 4.1) always being the best performing one. From here, more work is needed (a) to simulate data without relying on existing participants' parameter ranges, thus providing better insight into combinations of parameter values that might lead to better model recovery, (b) to identify other approaches for modelling the empathy learning task data (e.g., Bayesian learning models for belief updating, Kuzmanovic & Rigoux, 2017; Leong & Zaki, 2017), (c) to design alternative versions of the empathy learning task that more closely correspond to well-established computational models in the literature (e.g., a task where participants learn how to respond to outgroup members' outcomes by directly receiving or observing others receive social rewards and consequences – for example, verbal praise or criticism – from relevant others), and ultimately, (d) to design models that capture a wider array of competing learning hypotheses to ascertain whether there might be multiple learning processes that can give rise to the patterns of inaccuracy observed in real participants' perceptions of intergroup empathic norms.

In summary, the influence of ingroup and outgroup intergroup empathic norms on displays of intergroup empathy has been exclusively studied in isolation, making it difficult to predict how people would respond when both empathic norms are available

simultaneously and are either aligned or misaligned. Experiment 7 directly contrasted ingroup and outgroup intergroup empathic norms and showed that people are more easily motivated to display intergroup empathy in both public and private if they witness outgroup members expressing empathy for one's ingroup than if they witness ingroup members expressing empathy for the same outgroup. This result has promising practical implications for the improvement of intergroup relations as it renders outgroup intergroup empathic norms a particularly important target of empathy-building interventions. Nonetheless, people form perceptions of intergroup empathic norms in a biased manner such that exposure to even the most impartial intergroup interactions does not guarantee that these interactions will be represented as such.

General discussion

In this chapter, I have presented the findings from two experiments which investigated (a) whether outgroup intergroup empathic norms encourage intergroup empathy through reducing how much outgroup members are perceived as deserving of their adversities (Experiment 6b) and (b) how people's displays of intergroup empathy vary across contexts where multiple intergroup empathic norms are salient simultaneously and are either aligned or misaligned (Experiment 7). Although both experiments reverted to examining the causal influence of intergroup empathic norms on one's own displays of intergroup empathy, given that Experiment 7 did not directly pick up on the findings of Experiment 6b, I decided to provide extensive experiment-specific discussions as part of their respective sections. The general discussion will instead act as a conclusion, allowing me to comment on the broader implications of this chapter's experiments and outline avenues for future research.

First, Experiment 6b, with all its limitations, made the first step towards bridging two thus far independent strands of research. The first strand of research has been devoted to investigating how empathy is shaped by different beliefs about its nature (e.g., beliefs about

empathy's effortfulness, malleability or utility – Cameron et al., 2019; Hasson et al., 2022; Santos et al., 2022; Schumann et al., 2014) and cognitive appraisals of others' emotional circumstances (e.g., appraisals of deservingness, responsibility and behavioural stability – Betancourt, 1990; Davis & Gold, 2011; Decety et al., 2010; Vescio et al., 2003). The second strand of research has looked at how social norms influence empathy, without offering much insight into whether they also shape cognitive appraisals of others' emotions and empathy-related beliefs (e.g., Gubler et al., 2015; Hein et al., 2015; Nadler & Liviatan, 2006; Nook et al., 2016; Saguy et al., 2015; Tarrant et al., 2009). While past work has successfully isolated multiple factors with a potential to amplify or reduce empathy, future research should seek to address outstanding questions at the intersection of these factors. For example, future research could ask whether witnessing other people's displays of empathy influences own displays of empathy by increasing how grave and undeserving empathic targets' misfortunes seem, or by changing how effortful, malleable, or useful people perceive empathy overall.

Second, Experiment 7 shed light on how learning biases might contribute to the negativity and inaccuracy of intergroup empathic norm perceptions even in the absence of any exposure to unrepresentative intergroup content. Therefore, with Experiment 7, I join a growing number of researchers seeking to understand how learning operates in intergroup environments (Lindström et al., 2014; Over & McCall, 2018; Montrey & Shultz, 2022). Despite the focus on intergroup empathy, the findings of Experiment 7 call for a greater acknowledgement of the role of learning biases across a variety of intergroup phenomena such as false polarisation (i.e., inaccurate beliefs about the magnitude of political polarisation, Levendusky & Malhotra, 2016) and group meta-perceptions (i.e., inaccurate beliefs about how one's ingroup is perceived by outgroups, Lees & Cikara, 2020). In the context where multiple explanations of false polarisation and group meta-perceptions have been provided (including factors such as categorical thinking and emotional amplification,

Fernbach & Van Boven, 2022; Lees & Cikara, 2021), future research should aim to outline how different contributors are expected to manifest in relevant learning episodes (e.g., when people are exposed to a political outgroup member's opinion on a social issue). In addition to its theoretical implications, Experiment 7 offers several methodological recommendations relevant for the study of learning. First, while developmental psychology (Brkić et al., 2022; Räsänen et al., 2009) and behavioural economics (Thielmann et al., 2021) are no strangers to gamification, Experiment 7 has illustrated how gamified tasks can be designed to mimic observational social learning (i.e., learning by observing others) and capture theoretically relevant features of intergroup interactions. Second, Experiment 7 has also emphasised some of the common challenges associated with designing novel learning tasks and fitting computational models (e.g., parameter and model recovery) as well as some approaches that can be taken to mitigate these challenges (e.g., data simulation, adoption of model independent measures, Wilson & Collins, 2019). Overall, given the growing recognition that computational modelling provides a framework to delineate learning mechanisms that underlie social judgments and behaviour (see Cushman & Gershman, 2019; Hackel & Amodio, 2018; FeldmanHall & Nassar, 2021), computational models in combination with gamified learning tasks, akin to the one used in Experiment 7, can be expected to feature in an increasing number of scientific endeavours aimed towards understanding groups and group relations.

Third, both Experiments 6b and 7 sought to be mindful of the impact specific features of the experimental designs and stimuli as well as examined groups, group relations and environments could have had on results and conclusions. As such, these experiments transparently discussed not only which methodological shortcomings future studies would benefit from addressing, but also how broadly reported results are expected to generalise and what individual and situational factors are likely to act as moderators (Simons et al., 2017).

Overall, this granular discussion aims to motivate researchers to interpret current (and any future) findings less in light of specific group categories they are based on (e.g., minimal groups) and more in light of a variety of relevant features that were present in the examined intergroup context (e.g., absence of punishment for norm violation, visibility of one's empathic displays, interdependence). This approach reflects the belief that future research may be better served by tilting its focus in favour of using *contextual features* over *group categories* as units of explanation (see Cikara et al., 2022).

Finally, Experiment 7 represents an important continuation of Chapter 3. As a reminder, Chapter 3 examined how people's representations of known intergroup relations might inform their assumptions regarding intergroup empathy, but also theorised that the relationship between representations of intergroup relations and displays of intergroup empathy might be bidirectional. In particular, others' displays of intergroup empathy might provide one type of input based on which people can infer general characteristics of intergroup relations (e.g., their valence), with these general characteristics then used to predict intergroup feelings and behaviours in novel circumstances (e.g., if people assume that intergroup relations are negative, they might predict that one group will not display empathy in response to other group's outcomes). While experiments in Chapter 3 focused on the latter causal arrow (i.e., how representations of intergroup relations' valence inform predictions of intergroup empathy), Experiment 7 focused on the former (i.e., how displays of intergroup empathy inform representations of intergroup relations' valence). Importantly, in this thesis, the focus is not on the dimension of valence *per se*, but this dimension is used as a case study to illustrate the importance of understanding how people represent group relations in addition to groups' traits and how they combine both types of representations to make social predictions in intergroup contexts. Currently, social psychology is rife with low-dimensional models of groups' traits (e.g., Stereotype Content Model – Cuddy et al., 2009; Dual

Perspective Model – Abele & Wojciszke, 2014; Agency-Beliefs-Communion Model – Koch et al., 2016), but the domain of group relations has remained almost entirely unmapped and thus requires future scientific attention. In brief, psychological research should not neglect the fact that "…people must also represent, attend to, and think about the properties of the relationships between individuals [and groups]. Social cognition is […] concerned with both the nodes and the links in the social network" (Haslam & Fiske, 2004, p. 165).

In conclusion, in this chapter, I have sought to understand the mechanisms through which intergroup empathic norms influence displays of intergroup empathy (Experiment 6b), the preferences people display when faced with multiple irreconcilable intergroup empathic norms and the learning processes which underlie the formation of intergroup empathic norm perceptions (Experiment 7). Based on current results, portraying intergroup empathy as normative remains a promising strategy to encourage displays of empathy across group lines. Nonetheless, any psychological intervention aiming to influence intergroup empathic norms should be prepared to grapple with people's tendency to form biased perceptions of these norms even on the basis of the most impartial intergroup interactions.

Open practices

Experiment 7 was pre-registered, and the pre-registration can be found at this link: https://aspredicted.org/blind.php?x=9MH_W1W.

CHAPTER 5: GENERAL DISCUSSION

General overview of the chapter

In this thesis, I presented a series of experiments which investigated how group category membership and induced or perceived group relations influence two socio-affective processes: (a) judgments of facial expressions of emotion (Chapter 2) and (b) empathy (Chapters 3 and 4). Specifically, in the domain of facial expressions of emotion, Chapter 2 examined how minimal group membership of, and coalitional relations, with the expresser influence judgments of facial expressions. Furthermore, Chapter 2 investigated whether these factors exert their influence by altering patterns of visual attention, and whether peripheral physiology holds the potential to uncover steps in the processing of social information which might not always manifest in explicit judgments. In the domain of empathy, Chapter 3 provided a description of people's subjective perceptions of intergroup empathic norms, the accuracy of these perceptions, and their dependency on contextual information about empathic targets and their emotional experiences as well as on individual differences in how relevant group relations are perceived. Finally, Chapter 4 sought to provide insights on how intergroup empathic norms encourage empathy across group lines and which intergroup empathic norms are more likely to be followed in environments where both ingroup and outgroup intergroup empathic norms are active and promote irreconcilable emotional reactions (i.e., intergroup apathy versus intergroup empathy).

Across all empirical chapters, I strived to move beyond category-oriented interpretations, which focus on real-world or minimal group categories (i.e., social identities) as units of description and interpretation, and to embrace feature-oriented interpretations, which discuss relevant features of real-world or minimal ingroups and outgroups, their relations, and the environments in which they operate. Below, I summarise the key findings

of the research presented in this thesis, situate these findings in relation to the wider literature, and identify the most promising avenues for future research. I end by discussing broader theoretical considerations and illustrating ways in which future research can approach the study of "groups" and particularly, the study of their interplay with socio-affective processes.

The influence of group category membership and coalitional relations on judgments of facial expressions of emotion

Experiments 1-3 in Chapter 2 investigated the outstanding question of whether minimal ingroup biases in valence judgments of facial expressions can be better understood as a product of co-occurring coalitional relations (competition or cooperation), group membership (whatever remains of it once coalitional relations are taken out) or their interaction. This work was primarily motivated by the fact that previous research either experimentally confounded group membership with coalitional relations (Lazerus et al., 2016) or failed to account for the fact that these two factors are implicitly confounded (Dunham, 2011; Gamond et al., 2017) as people automatically anticipate cooperation with minimal ingroup members, even in the absence of any explicit instructions (Balliet et al., 2014; Brewer, 2008; Misch et al., 2021; Yamagishi & Kiyonari, 2000).

Findings

Experiments 1 & 2. Experiments 1 and 2 in Chapter 2 showed that valence judgments of angry, ambiguous, and happy facial expressions of emotion are affected by both minimal group membership of the expresser (i.e., whether the expresser is a minimal ingroup or a minimal outgroup member) and coalitional relations with the expresser (i.e., whether one expects to cooperate or compete with the expresser), while it found no evidence that these two factors influenced arousal and genuineness judgments of the same facial expressions. To unpack these findings, minimal ingroup members' facial expressions were always judged as communicating more positive emotions compared with minimal outgroup members' facial expressions (i.e., *ingroup positivity bias*). Additionally, coalitional relations influenced valence judgments with participants reporting that facial expressions of those with whom they expected to cooperate conveyed more positive emotions than facial expressions of those with whom they expected to compete (i.e., *cooperation positivity bias*). These results

provided the first evidence that previously documented ingroup positivity bias in judgments of facial expressions (Lazerus et al., 2016) cannot be accounted for by underlying coalitional relations.

Experiment 3. Experiment 3 expanded on the findings from Experiments 1 and 2 to assess whether minimal group membership of, and coalitional relations with, the expresser exert their influence on judgments of facial expressions by changing how people allocate attention to expressers' facial features. Specifically, Experiment 3 started with three hypotheses. First, it addressed the possibility that participants process minimal ingroup members' and cooperative others' facial expressions more holistically than minimal outgroup members' and competitive others' facial expressions (Hugenberg & Corneille, 2009; Michel et al., 2007). Second, it addressed the possibility that participants engage in greater attentional search for signs of positivity in minimal ingroup members' and cooperative others' facial expressions than in minimal outgroup members' and competitive others' facial expressions[79]. Finally, Experiment 3 remained open to the possibility that minimal group membership and coalitional relations do not change how participants attend to facial features, but rather exert their influence only on valence judgments of otherwise similarly explored facial expressions (Lazerus et al., 2016). Experiment 3 found no evidence that the expresser's minimal group membership and coalitional relations with them alter patterns of visual attention to facial features. Specifically, no differences were found in the average duration of individual fixations on different facial features (i.e., eyes, nose, mouth), total viewing time of these facial features nor the number of intra- and inter-featural saccades. These results suggest that minimal group membership of, and coalitional relations with, the expresser are

[79] More details on the exact indices of holistic processing and attentional search for signs of positivity can be found in Chapter 2.

unlikely to influence valence judgments of facial expressions by triggering top-down attentional shifts.

Finally, Experiment 3 employed measures of peripheral physiology – pupil dilation and skin conductance response (SCR) – to test whether these measures might reveal implicit processing steps (i.e., participants' sensitivity to incongruent social information about the expresser) that did not necessarily manifest in participants' judgments of facial expressions. Experiment 3 hypothesised that participants might harbour ingroup-cooperation and outgroup-competition associations (and similar group-emotion and coalition-emotion associations) and that they might show increased pupil dilation and SCR when these associations are violated; in particular, when they are introduced to ingroup members with whom they are to compete and outgroup members with whom they are to cooperate (see Hepach & Westermann, 2013).

Experiment 3 found support for the idea that people's processing of facial expressions might be characterised by sensitivity to violations of group-coalition associations. Specifically, participants showed greater pupil dilation when they were viewing facial expressions of "incongruent" expressers (i.e., ingroup members with whom they were to compete and outgroup members with whom they were to cooperate) relative to "congruent" expressers (i.e., ingroup members with whom they were to cooperate and outgroup members with whom they were to compete). SCR did not reveal the same two-way interaction, but rather suggested greater physiological responsivity to facial expressions of minimal ingroup members and competitive others versus facial expressions of minimal outgroup members and cooperative others. While the two measures might not be entirely aligned, they agree that participants showed increased physiological responsivity to minimal ingroup members with whom they were to compete, which could signal that participants were more sensitive to violations of intragroup than intergroup coalitional expectations.

In addition to these results, participants also displayed greater physiological responsivity, as indexed by SCR but not pupil dilation, in response to textual cues informing them that they were to compete versus cooperate with the upcoming expresser. This finding expands on previous literature which suggested that competition puts people in a state of high arousal (Scott & Cherrington, 1974; Ku et al., 2005; To et al., 2018) to indicate that even mere *anticipation* of competition can achieve the same effect. Finally, participants displayed the greatest physiological responsivity, as indexed by pupil dilation but not SCR, when viewing angry facial expressions followed by ambiguous and happy facial expressions. This finding adds to past work which suggested that facial expressions of anger might increase perceivers' physiological arousal because they signal interpersonal threat (Kret et al., 2013).

Implications and future directions

The implications of results from Experiments 1-3 are threefold. First, current results are relevant for coalitional accounts of the minimal group paradigm, which suggest that (assumed) coalitional relations are the ingredient which imbues minimal ingroups with affiliative meaning and underlies minimal group biases documented in the literature (Cikara, 2020; Misch et al., 2021; Pietraszewski, 2013; Yamagishi & Kiyonari, 2000; see also Powell, 2022). In line with this view, it has been shown that once coalitional expectations stemming from minimal group membership are violated (e.g., people no longer expect to cooperate with minimal ingroup but rather minimal outgroup members), minimal group preferences diminish or completely disappear (Locksley et al., 1980; Misch et al., 2021; Rabbie et al., 1989; Yamagishi & Kiyonari, 2000). If coalitional relations similarly accounted for minimal group biases in valence judgments of facial expressions, one would expect to observe only the effect of coalitional relations, but not the effect of minimal group membership.

This is, however, not what Experiments 1-3 found. Specifically, minimal group membership continued to influence valence judgments of facial expressions even after it was

decoupled from coalitional relations such that minimal ingroup members were not a reliable cue for cooperation and minimal outgroup members were not a reliable cue for competition. These results suggest that previously documented ingroup positivity bias in judgments of facial expressions (Lazerus et al., 2016) was likely amplified but not entirely driven by co-occurring coalitional relations. Overall, based on current results, it seems that minimal group, and coalition effects on valence judgments of facial expressions could be better understood as specific examples of *affective priming*, where the evaluative response to expectations of shared group membership and cooperative relations with the expresser translates to the experience and the final valence judgments of facial expressions (Fazio, 2001; Houwer, 2009).

As discussed in Chapter 2, it is worth noting that participants in Experiments 1-3 were led to believe that minimal ingroup and outgroup members did not choose to compete or cooperate but were randomly assigned to a competitive or a cooperative context. This cover story could have prevented potentially negative evaluations of minimal ingroup members with whom participants were to compete. Future research should thus investigate the generalisability of these findings to minimal ingroup members who choose to deviate from the expectation to cooperate instead of simply being allocated to a competitive intragroup context. Given that those who deliberately violate the norm of intragroup cooperation tend to be negatively evaluated (Mendoza et al., 2014; Wang et al., 2016), if valence judgments of facial expressions indeed depend on evaluations triggered by the expresser, facial expressions of minimal ingroup members who choose to compete against one's own group should be judged as conveying more negative emotions than facial expressions of minimal ingroup members who were assigned to compete against one's own group.

Second, to the best of my knowledge, Experiment 3 is the first study to investigate whether visual attention to facial expressions of emotion is impacted by minimal group

membership of, and coalitional relations with, the expresser. To recap, Experiment 3 showed that minimal group membership of, and coalitional relations with, the expresser are unlikely to exert their influence on valence judgments of facial expressions by changing how perceivers allocate visual attention to different features of facial expressions. This result is in line with prior research which examined face (i.e., identity) processing and showed that although minimal ingroup members' faces capture more visual attention overall when presented alongside minimal outgroup members' faces, how visual attention is distributed once it is within minimal ingroup members' faces or minimal outgroup members' faces is identical (Kawakami et al., 2014; Correll & Hudson, 2020). Although the findings of Experiment 3 say little about whether minimal group membership and coalitional relations might affect visual attention under different conditions (e.g., in face-to-face interactions, in the presence of different goals), they do suggest that minimal group biases in judgments of facial expressions, documented using similar experimental paradigms and stimuli as the ones adopted in Experiment 3, are unlikely to have resulted from specific visual attention strategies (Dunham, 2011; Lazerus et al., 2016; Young & Hugenberg, 2010).

Nevertheless, given how sensitive visual attention is to a variety of bottom-up (e.g., salient features, dynamics of expression development, expresser's gaze direction, Juarez et al., 2019) and top-down factors (e.g., one's goals in a social interaction, one's knowledge of others, Hessels, 2020), there is likely no universal visual attention strategy reserved for (minimal or real-world) ingroups versus outgroups, and cooperative others versus competitive others which is activated regardless of the stimulus (e.g., facial expression) and task (e.g., whether perceiver is evaluating valence of the facial expression or gender of the expresser) at hand. Therefore, instead of searching for universal principles (e.g., increased attention to the eyes, Kawakami et al., 2014), future research should embrace the situational variance of visual attention and aim to identify circumstances under which group

membership and coalitional relations might become particularly relevant in guiding not only how social information is visually sampled, but also how social information is signalled to others (i.e., *dual function of gaze*, Gobel et al., 2015).

Third, Experiment 3 illustrated how physiological responses – pupil dilation and SCR – can be used to provide important insights which could not be gleaned from participants' judgments of facial expressions alone. Specifically, pupil dilation and SCR revealed sensitivity to incongruence between minimal group membership and coalitional relations, providing further support to the idea that people have a generalised expectation of within-group cooperation (Balliet et al., 2014; Brewer, 2008; Misch et al., 2021; Yamagishi & Kiyonari, 2000), and to a lesser extent a generalised expectation of between-group competition. From here, future research should examine whether physiological responses to incongruent group-coalition pairings vary depending on the person's idiosyncratic endorsement of congruent group-coalition pairings. For instance, do people who more strongly associate (minimal) ingroup members with cooperation show stronger physiological responses once this association is violated? If this is confirmed, it would suggest that physiological responses might offer insight into individuals' beliefs about how different social phenomena, even beyond groups and coalitions, correlate (see Stolier & Freeman, 2016).

Perceptions of empathic norms in intergroup contexts

Experiments 4-6a in Chapter 3 sought to describe people's unaltered perceptions of empathic norms in intergroup contexts in contrast to most previous research which has manipulated these norms to examine their influence on one's own intergroup empathy or related outcomes (e.g., intergroup dehumanisation and prejudice, Gubler et al., 2015; Nook et al., 2016; Saguy et al., 2015; Tarrant et al., 2009, but see Penić et al., 2021). Specifically, Chapter 3 recognised that in real life people are rarely, if ever, exposed to one-off summary

information about intergroup empathic norms, as participants are in psychological experiments. However, this does not prevent people from forming subjective perceptions of intergroup empathic norms on the basis of broad personal experience (e.g., self-experienced intergroup interactions, observations of others). Given that norm perceptions have a greater influence on behaviour than actual norms and even personal beliefs (Jachimowicz et al., 2018; Tankard & Paluck, 2016; Vallentin-Holbech et al., 2017), it is of crucial importance to characterise perceptions of intergroup empathic norms as they represent the very normative context in which people decide to extend their empathy (or not) across group lines.

Findings

Experiment 4. In the post-conflict context of Croatia and Serbia, Experiment 4 showed that perceptions of intergroup empathic norms tend to be overly negative and inaccurate. Specifically, participants assumed that ethnic ingroup and outgroup members would feel more intergroup empathy bias than (a) actual ingroup and outgroup members in fact did, and than (b) participants reported feeling themselves. Even though Experiment 4 focused on perceptions of intergroup empathic norms, it should be noted that the same pattern of results was observed when comparing how much participants discriminated on the basis of ethnic group membership in the way they *perceived* empathic targets' emotions and how much they assumed ingroup and outgroup members discriminate in the way they *perceive* empathic targets' emotions. Second, Experiment 4 documented *empathic norm perception asymmetry*, wherein ethnic outgroup members were assumed to discriminate more than ethnic ingroup members in their empathy, but not in the way they perceive emotions. Third, participants' perceptions of intergroup empathic norms were sensitive to the availability of contextual information about empathic targets and their emotional experiences. Specifically, when participants reported their perceptions in relation to known empathic targets and their specific emotional experiences, they expected ethnic ingroup and outgroup members to show

less intergroup empathy bias than when participants reported their perceptions in the abstract, without known empathic targets and specific emotional experiences. Finally, participants' perceptions of intra- and intergroup relations' valence (e.g., how positively participants thought ethnic outgroup members feel towards ethnic ingroup members) were crucial in predicting corresponding perceptions of empathic, but not emotion-perception, norms. For instance, the more negative feelings participants thought that ethnic outgroup members had towards the ethnic ingroup, the less intergroup empathy they assumed ethnic outgroup members would feel. In sum, Experiment 4 outlined that perceptions of intergroup empathic norms tend to be overly negative and inaccurate, particularly (a) when there are no known empathic targets with specific emotional experiences, and (b) among those who perceive relevant intergroup relations in a negative light.

Experiment 5. Experiment 5 sought to replicate the findings of Experiment 4 in the zero-sum, political context of Brexit and to examine the extent to which perceptions of intergroup empathic norms vary as a function of how much empathic targets' emotional experiences make their social (i.e., political) identities salient. First, as in Experiment 4, participants assumed that political ingroup and outgroup members were more biased in their empathy (and emotion perception) than participants were themselves.

Second, expanding on Experiment 4, Experiment 5 showed that perceptions of intergroup empathic (and emotion-perception) norms were more likely to be overly negative and inaccurate when empathic targets' emotional experiences made their social (i.e., political) identities salient (e.g., by revealing how empathic targets came to adopt their social identity) than when empathic targets' emotional experiences drove attention away from their social identities (e.g., by pointing to empathic targets' specific life circumstances that have nothing to do with their social identities). Even though I emphasised increased salience of empathic targets' social identities in the interpretation of this finding, in Chapter 3, I

speculated that participants' own empathy and perceptions of empathic norms could have also been influenced by judgments of how justifiable it was for an empathic target to attribute their emotional experience to Brexit-related concerns (e.g., how justified participants thought it was for an empathic target to attribute their experience of losing a job in the tourism industry to Brexit-related concerns).

Third, Experiment 5 documented empathic norm perception asymmetry such that political outgroup members were expected to discriminate more than political ingroup members in their empathy when empathic targets' social identities were salient. However, there was no evidence of such asymmetry when emotional experiences reduced the salience of empathic targets' social identities. Finally, in line with Experiment 4, participants' perceptions of intra- and intergroup relations' valence predicted participants' perceptions of empathic, but not emotion-perception, norms in contexts where it was relevant to think of empathy or a lack thereof as a product of group relations (i.e., when emotional experiences facilitated construal of empathic targets as group members rather than individuals). In sum, Experiment 5 replicated findings from Experiment 4 and emphasised that perceptions of intergroup empathic norms are likely to be particularly negative in the very context where their positivity would bring the most benefit – when people learn about empathic targets' emotional experiences which are intertwined with their social (i.e., political) identities.

Experiment 6a. Finally, Experiment 6a showed that even a minimal intergroup context can produce assumptions of others' intergroup empathy bias, but this assumed intergroup empathy bias was significantly smaller than that reported by participants in Experiments 4 and 5. Furthermore, Experiment 6a found no evidence of empathic norm perception asymmetry, suggesting that its occurrence might be associated with features shared by the two real-world intergroup contexts examined in Experiments 4 and 5 (e.g., history of conflict, competition).

Implications and future directions

The implications of Chapter 3 can be summarised in three points. First, Chapter 3 expands on previous research which has found that people overestimate both (political) ingroup and outgroup members' prejudice towards and dehumanisation of each other but are more inaccurate in their (over)estimates of (political) outgroup members' than ingroup members' antagonism (Lees & Cikara, 2020; Moore-Berg et al., 2020; Ruggeri et al., 2021). Chapter 3 showed that these overestimation tendencies do not only characterise individuals' assumptions about general intergroup prejudice and dehumanisation, but also their assumptions about intergroup empathy. Specifically, participants in Experiments 4 and 5 similarly overestimated ethnic and political ingroup and outgroup members' intergroup empathy bias, with this tendency only reduced, but not completely eliminated, when people reported their perceptions of empathic norms in the context of single empathic targets and their individuating emotional experiences. Furthermore, in line with past research, Chapter 3 documented that in competitive intergroup contexts people harbour more negative and inaccurate perceptions of outgroup intergroup empathic norms (i.e., assumptions about how much empathy outgroup members feel for the ingroup) than of ingroup intergroup empathic norms (i.e., assumptions about how much empathy ingroup members feel for the outgroup). Importantly, Chapter 3 did not find this *empathic norm perception asymmetry* in a minimal intergroup context, which suggests that its occurrence might be propelled by the presence of competition and conflict. In summary, current findings provide support to the idea individuals' reasoning in *competitive* intergroup contexts can be summarised as "we might dislike/dehumanise/fail to empathise with them, but they dislike/dehumanise/fail to empathise with us more". As discussed in Chapter 3, it remains for future research to examine whether own-group love/other-group hate motive attribution asymmetry (Waytz et al., 2014) might be behind this empathic norm perception asymmetry. Specifically, it is

possible that the ingroup's assumed intergroup prejudice, dehumanisation and/or empathy bias are seen as arising from own-group love, but not other-group hate, whereas the competitive outgroup's assumed intergroup prejudice, dehumanisation and/or empathy bias are seen as a product of both own-group love and other-group hate.

Second, the findings from Chapter 3 also bear practical relevance to future empathy-building interventions aiming to change people's perceptions of intergroup empathic norms. Specifically, they provide valuable insights into the pattern of pre-existing beliefs about intergroup empathy that participants might have before entering empathy-building interventions, allowing these interventions to tailor their content to achieve maximal plausibility and effectiveness (Tankard & Paluck, 2016). Nonetheless, researchers, practitioners and/or policymakers seeking to change perceptions of intergroup empathic norms in contexts which do not share crucial features with the contexts examined in Chapter 3 (e.g., history of conflict, competition) should always investigate perceptions of relevant intergroup empathic norms prior to deploying empathy-building interventions.

Third, individuals' perceptions of the valence of intergroup relations, embedded in perceptions of intergroup prejudice, are not independent of perceptions of intergroup empathy. The more people believed that the groups in question had negative feelings towards each other, the less they expected them to have empathy for each other's misfortunes. One aspect of this result is important to emphasise. Specifically, how much empathy individuals assumed outgroup members felt towards the ingroup (and vice versa) did not depend on how positive and empathetic outgroup (or ingroup) members were perceived to be as such, but rather on how the *relations* between the two groups were perceived. In other words, even if members of hypothetical groups A and B are stereotypically perceived as warm and friendly, what crucially matters when predicting how members of group A will behave towards members of group B is not group A's or group B's general warmth, but the warmth of the

relations between them. In essence, this finding emphasises the importance of considering how people make sense of *relations* in addition to *traits* as well as how relational and trait knowledge are intertwined and jointly used to predict and adjust to specific intergroup (and interpersonal) interactions. Nevertheless, although social psychology has put a lot of effort into understanding relations between countless real-world groups (e.g., relations between political groups in the United States, relations between Greek Cypriots and Turkish Cypriots) as well as different types of relations between individuals (e.g., friendships, romantic relationships, child-parent relations), it has not yet managed to abstract dimensions that organise people's cognition of relational dynamics from these specific examples (for classical examples of relational taxonomies, see Fiske, 1992; Haslam & Fiske, 1999).

This leaves an abundance of outstanding questions for future research. First, future research should map the dimensions people rely on to represent the richness and complexity of social relations (for similar approaches which identified a low-dimensional structure of traits, mental states, and actions, see Tamir et al., 2016; Tamir & Thornton, 2018; Thornton & Tamir, 2022) and test to what extent the same low-dimensional framework replicates across different cultures, languages, and points in time (see Thornton et al., 2022). Second, future research should also test to what extent the same low-dimensional framework adequately captures both representations of interpersonal and intergroup relations (see Bruckmüller & Abele, 2013). Third, more work is needed to understand how people learn about social relations over the course of development (Thomas et al., 2022) and how keeping track of relations in one's social environment allows inferences about the existence of groups, even in the absence of labelable group categories (see Gershman & Cikara, 2020; Pietraszewski, 2022). Fourth, future research should compare how people think about their own personal and group relations with how they think about others' personal and group relations (see Thornton et al., 2019). Finally, more work should be devoted to examining

traits alongside social relations; for instance, how different traits are valued within different types of relations (Abele & Brack, 2013) and how different traits shape and are shaped by specific relations with others (Asendorpf & Wilpers, 1998). In sum, while social psychology has multiple, vastly adopted models to characterise individuals' and groups' traits (e.g., Stereotype Content Model – Cuddy et al., 2009; Dual Perspective Model – Abele & Wojciszke, 2014; Agency-Beliefs-Communion Model – Koch et al., 2016), it is in dire need of an empirically supported model to characterise social relations on both an interpersonal and an intergroup level.

Learning to empathise: How do intergroup empathic norms influence outgroup-oriented empathy?

Experiments 6b and 7 in Chapter 4 sought to provide insights into the mechanisms through which intergroup empathic norms achieve their influence and into the preferences people display when faced with ingroup and outgroup intergroup empathic norms that demand misaligned emotional reactions (e.g., intergroup apathy versus intergroup empathy). Experiment 6b expanded on previous research which has shown that if perceivers think that empathic targets deserve their negative outcomes, they tend to feel schadenfreude and if they think that empathic targets do not deserve their negative outcomes, they tend to feel empathy (Feather, 2006; Feather et al., 2013; Hudson, 2022). Therefore, Experiment 6b examined whether outgroup intergroup empathic norms increase intergroup empathy by reducing how deserving of their misfortunes outgroup empathic targets are thereafter perceived.

Experiment 7 expanded on previous research which found that *reciprocity* of outgroup members' behaviour enhances cooperation more than *conformity* to ingroup members' behaviour (Romano & Balliet, 2017, see also Chiang, 2021). In particular, to the best of my knowledge, Experiment 7 is the first study to expose participants to both ingroup and outgroup intergroup empathic norms and track whether those participants preferentially

conform to ingroup norms or reciprocate outgroup norms. Beyond comparing conformity and reciprocity, Experiment 7 assessed whether individually or contextually determined reputational concerns predict which empathic norms participants adhere to as well as whether participants' perceptions of intergroup empathic norms misrepresent the intergroup empathic norms they were in fact exposed to.

Findings

Experiment 6b. In the context of minimal groups, Experiment 6b was designed to assess whether one mechanism through which outgroup empathic norms influence own intergroup empathy involves changing how much outgroup empathic targets' misfortunes are subsequently perceived as deserved. On one hand, learning that impartial empathy is normative among outgroup members was expected to reduce the extent to which outgroup empathic targets' misfortunes trigger attributions of deservingness, resulting in increased empathy for outgroup empathic targets and correspondingly, reduced intergroup empathy bias. On the other hand, learning that biased empathy is normative among outgroup members was expected to increase the extent to which outgroup empathic targets' misfortunes trigger attributions of deservingness, resulting in reduced empathy for outgroup empathic targets and correspondingly, increased intergroup empathy bias.

Experiment 6b found no evidence in support of this hypothesis. In particular, rather than increasing empathy solely for outgroup empathic targets, impartial outgroup empathic norms were shown to enhance empathy for both ingroup and outgroup empathic targets only in comparison to the control group, where no outgroup empathic norms were illustrated. However, this general empathy increase did not stem from impartial outgroup empathic norms affecting deservingness appraisals. Nevertheless, due to several methodological shortcomings of Experiment 6b, which I discussed in Chapter 4 (e.g., implausible experimental manipulation, highly individuating and occasionally ambiguous stimuli), future

research should not abandon deservingness appraisals as a candidate mechanism on the basis of these non-significant results.

Experiment 7. In Experiment 7, participants completed a novel empathy learning task which was designed to investigate how displays of intergroup empathy varied depending on the exact combination of ingroup and outgroup intergroup empathic norms participants were exposed to. Furthermore, this empathy learning task allowed examination of how participants' displays of intra- and intergroup empathy varied between contexts where these displays were public, and thus open to reputational scrutiny, versus private, and thus protected from reputational scrutiny. Finally, given that Experiment 7 had full control over the empathic norms participants were exposed to, it was well-placed to assess whether participants' final perceptions of these norms deviated from reality. Of note, Experiment 7 was conducted in the context of minimal groups, where participants though they were interdependent with other minimal ingroup members, but not with minimal outgroup members.

In terms of results, Experiment 7 showed that participants primarily adjusted their intergroup empathy to match the emotional reactions outgroup members displayed towards the ingroup. Specifically, if outgroup members displayed intergroup empathy, participants not only reciprocated intergroup empathy at that particular time, but also subsequently reported more positive feelings towards and greater identification with the outgroup. Although statistical trends suggested that participants also considered ingroup intergroup empathic norms, all conformity-related effects fell short of statistical significance. In summary, reciprocity outperformed conformity not only in promoting immediate intergroup empathy, but also in promoting more positive intergroup feelings overall.

Contrasting participants' public and private displays of empathy, participants displayed the same degree of *intragroup* empathy regardless of whether their responses were

public or private, but they displayed less *intergroup* empathy when their responses were private as opposed to public. Although private intergroup empathy was reduced in comparison to public intergroup empathy regardless of the operative intergroup empathic norms, this reduction was the smallest in experimental conditions where outgroup members showed empathy for ingroup members' outcomes, thus underscoring the motivational potency of outgroup over ingroup intergroup empathic norms.

Experiment 7 also zoomed in on experimental conditions where ingroup and outgroup intergroup empathic norms were misaligned and provided evidence that displays of empathy which were a result of conformity to ingroup members' intergroup empathy reduced in the absence of reputational concerns more than displays of empathy which were a result of reciprocity of outgroup members' intergroup empathy. This result suggests that participants' conformity to ingroup members' intergroup empathy is driven by reputational concerns to a greater extent than participants' reciprocity of outgroup members' intergroup empathy is.

However, the same conclusion was not supported under an alternative analytical approach where participants' self-reported reputational concerns were used to predict which intergroup empathic norm they adhered to in misaligned experimental conditions. If reputational concerns were indeed behind participants' conformity, one would expect participants with high reputational concern (i.e., participants who were more concerned about how ingroup members evaluated them) to display higher intergroup empathy than participants with low reputational concern in the conform-to-empathise condition (i.e., ingroup empathy for the outgroup, outgroup apathy for the ingroup condition) while this pattern should be reversed in the reciprocate-to-empathise condition (i.e., ingroup apathy for the outgroup, outgroup empathy for the ingroup condition). This is, however, not what Experiment 7 documented. Specifically, participants with low reputational concern showed higher intergroup empathy regardless of whether intergroup empathy was normative or

counter-normative among ingroup members. In Chapter 4, I speculated that participants may have interpreted the self-report measure of reputational concerns differently in different experimental conditions and depending on the degree of conformity they displayed, and that this ambiguity could have contributed to the discrepancy in results. In summary, despite these inconsistencies, Experiment 7 provided some initial evidence that reputational concerns are more likely to drive conformity to ingroup intergroup empathic norms than reciprocity of outgroup intergroup empathic norms.

The final finding of Experiment 7 was that participants did not form perceptions of intergroup empathic norms in an unbiased manner, which meant that these perceptions were not an accurate reflection of the intergroup empathic norms that people were in fact exposed to. On one hand, participants in the experimental conditions where ingroup and outgroup intergroup empathic norms were aligned showed a tendency to underestimate how much intergroup empathy both ingroup and outgroup members displayed in the empathy learning task. This tendency could be recreated by simulating data under a model which assumed that participants have negative pre-existing beliefs about intergroup empathy (i.e., *negative priors*) and that they update their perceptions of intergroup empathic norms more on the basis of negative prediction errors (i.e., someone displaying less intergroup empathy than expected) than positive prediction errors (i.e., someone displaying more intergroup empathy than expected, *negativity bias*). On the other hand, participants in the experimental conditions where ingroup and outgroup intergroup empathic norms were misaligned showed a tendency to underestimate how much intergroup empathy the empathetic group displayed and to overestimate how much intergroup empathy the unempathetic group displayed in the empathy learning task. This tendency could be recreated by simulating data under a model which assumed that participants' learning of intergroup empathic norms was not only affected by *negative priors* and *negativity bias*, but also *reciprocity bias*, wherein participants

generalised one group's emotional reactions to the other group. In summary, these findings indicate that overly negative and inaccurate perceptions of intergroup empathic norms can arise as a product of biased learning, even when people are exposed to the most impartial intergroup interactions.

Implications and future directions

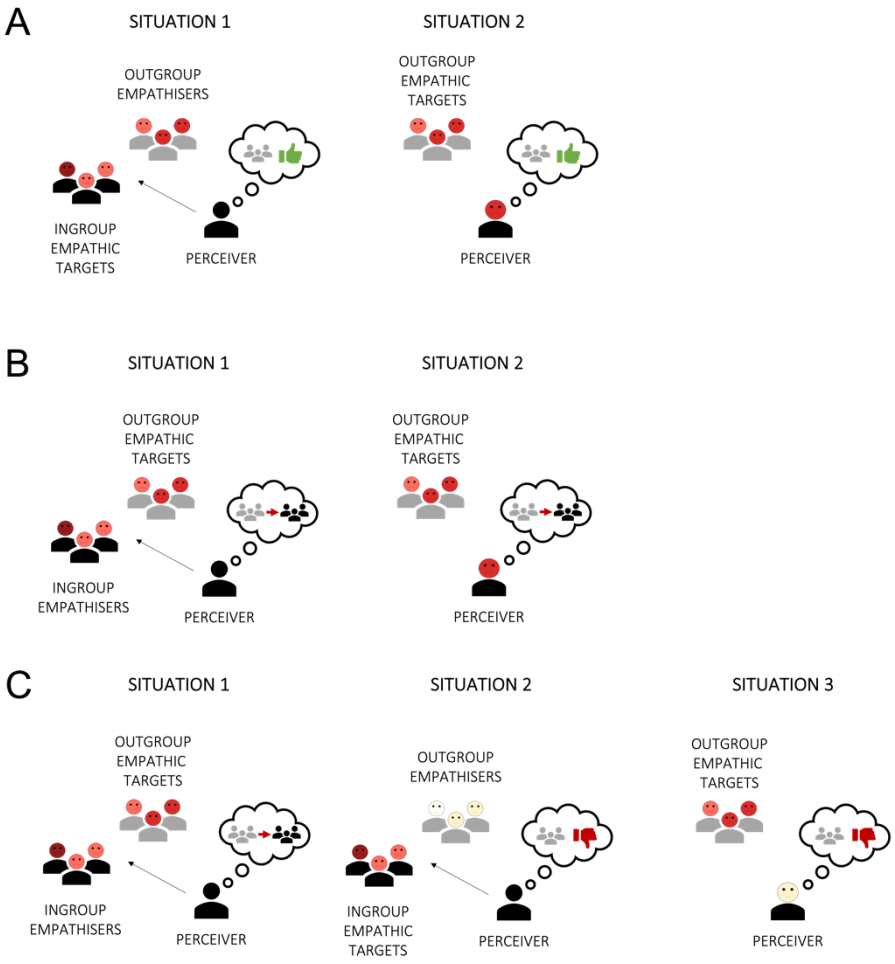
The implications of the results reported in Chapter 4 are threefold. First, with regard to prior research on empathy's selectivity depending on empathic targets' characteristics, studies have shown that if empathic targets are perceived as immoral and undeserving of empathy, people feel less empathy and more schadenfreude in response to their misfortunes (Cameron et al., 2022; Wang et al., 2022). Combining this past work with results from Experiment 7 provides a potential explanation of why reciprocity of outgroup intergroup empathic norms has a better chance of motivating intergroup empathy than conformity to ingroup intergroup empathic norms, across a variety of intergroup environments. Specifically, outgroup intergroup empathic norms might serve as a *direct* signal of whether outgroup members are deserving of empathy, while ingroup intergroup empathic norms only serve as an *indirect* signal of such deservingness due to people's generalised expectations of reciprocity.

On one hand, witnessing outgroup members displaying empathy for ingroup members' outcomes *directly* enhances perceptions of outgroup members' deservingness, encouraging perceivers to reciprocate intergroup empathy (Figure 5.1A). On the other hand, witnessing ingroup members displaying empathy for outgroup members' outcomes only *indirectly* encourages intergroup empathy because perceivers expect outgroup members to reciprocate such empathy when given the chance (Figure 5.1B). If this expectation of outgroup members' reciprocity is violated, ingroup intergroup empathic norms lose most of their influence on intergroup empathy (see Figure 5.1C). Of course, in the complexity of real-

world intergroup contexts, people are likely to engage in additional layers of motivated reasoning, potentially discounting outgroup members' empathy as inauthentic and/or manipulative (Shnabel et al., 2015). Nevertheless, perceptions of outgroup members' morality and deservingness of empathy should constitute a focal point of future research. Finally, linking approaches taken in Experiments 6b and 7, future research should examine whether changing perceptions of outgroup members' deservingness of empathy exerts its influence by altering appraisals of how much outgroup members deserved their specific emotional experiences. In other words, if perceivers believe that outgroup members are worthy of empathy overall, will this belief encourage empathy by altering how deserved perceivers judge outgroup members' specific misfortunes to be?

Figure 5.1

Overview of how ingroup and outgroup intergroup empathic norms might signal outgroup members' deservingness of empathy



Note. (A) Outgroup intergroup empathic norms acting as a *direct* signal that outgroup members are deserving of empathy. In situation 1, a perceiver witnesses outgroup members (i.e., empathisers) displaying empathy for ingroup members (i.e., empathic targets), based on which they conclude that outgroup members are moral and deserving of reciprocal empathy, as indicated by the thumbs-up sign in the thought bubble. In situation 2, the perceiver shows empathy for outgroup members given outgroup members' deservingness of such response. (B) Ingroup intergroup empathic norms acting as an *indirect* signal that outgroup members are deserving of empathy. In situation 1, a perceiver witnesses ingroup members (i.e., empathisers) displaying empathy for outgroup members (i.e., empathic targets), based on which they conclude that outgroup members are likely to reciprocate such empathy if given the chance, as indicated by the thought bubble, and are thus moral and

deserving of reciprocal empathy. In situation 2, the perceiver shows empathy for outgroup members given outgroup members' deservingness of such response. Importantly, in situation 2, perceiver's empathy for outgroup members depends on the expectation of reciprocity, but outgroup members did not have the opportunity to confirm or disconfirm this expectation. (C) Ingroup intergroup empathic norms acting as an *indirect* signal that outgroup members are deserving of empathy, with outgroup members subsequently violating expectation of reciprocity. Same as in scenario B, in situation 1, a perceiver witnesses ingroup members (i.e., empathisers) displaying empathy for outgroup members (i.e., empathic targets), based on which they conclude that outgroup members are likely to reciprocate such empathy if given the chance, as indicated by the thought bubble, and are thus moral and deserving of reciprocal empathy. In situation 2, a perceiver witnesses outgroup members displaying apathy as opposed to empathy for ingroup members, thus violating expectation of reciprocity. Due to this violation, the perceiver concludes that outgroup members are no longer deserving of reciprocal empathy, as indicated by the thumbs-down sign in the thought bubble. In response, in situation 3, the perceiver does not show empathy for outgroup members. (A-C) Situations which include a diagonal arrow are the situations in which the perceiver acts as a bystander, and situations which do not include this arrow are the situations in which the perceiver adopts the role of the empathiser. Furthermore, in all sub-figures, empathy is assumed if there is valence matching between empathic targets' emotions and empathisers' emotions (e.g., if empathic targets display negative emotions and empathisers respond with negative emotions).

Second, the very fact that Experiment 7 documented that own displays of intergroup empathy can be quickly encouraged in the presence of outgroup members' intergroup empathy supports the view that intergroup empathy is more a matter of motivation than ability (Cameron et al., 2017; Weisz et al., 2021; Zaki, 2014). However, the current experimental design also leaves many questions unanswered and thus raises a number of questions for future research to address. One such question concerns the homogeneity of group members' emotional reactions in Experiment 7. Specifically, one feature of Experiment 7 was that different members of the same group were governed by the same algorithm in the empathy learning task, resulting in them displaying similar levels of intergroup empathy. This makes it difficult to generalise the present results to more heterogeneous environments where different group members display different levels of

intergroup empathy. Future studies may thus expand on the current paradigm and manipulate the heterogeneity of group members' behaviour to examine how increased heterogeneity affects people's perception of and adaptation to intergroup empathic norms. Given that people have been shown to infer the moral status of a certain behaviour (i.e., whether the behaviour is right or wrong) based on its relative frequency (Lindström et al., 2017), it is likely that intergroup empathy needs to be widely displayed by group members in order for it to be perceived as normative and as a right course of action.

Another feature of Experiment 7 was that participants displayed emotional reactions towards the same outgroup members whose intergroup apathy or intergroup empathy they witnessed. Therefore, the results from Experiment 7 tell us little about whether participants would generalise their learnt empathic reactions to novel outgroup members, whom they had never previously seen. To address this outstanding question, future research could have participants learn about outgroup intergroup empathic norms from one group of outgroup members, and then have them proceed to a test phase where they would emotionally respond to outcomes experienced by new outgroup members (see Hackel et al., 2022 for an example). Finally, future research should explore how people's experience with one norm affects their subsequent adaptation to another norm (Hertz, 2021). For instance, if people first learn that intergroup apathy is normative among outgroup members, how will that affect their behaviour once the outgroup norm changes to intergroup empathy? Similarly, will all positive effects of witnessing intergroup empathy among outgroup members be lost if people are introduced to new outgroup members (or even a single outgroup member) who do not similarly display intergroup empathy? Overall, answering all these questions should lead to a better understanding of the conditions necessary for short-term adaptation to intergroup empathic norms to evolve into a long-lasting tendency.

Third, as discussed in Chapter 4, Experiment 7 provided important initial insights into one of the ways in which perceptions of empathic norms (and many other intergroup perceptions) may be formed – through observing individual ingroup and outgroup members' emotional reactions towards each other's outcomes – and suggested how intergroup processes are likely to interface with observational learning. Specifically, by comparing the final product of learning (i.e., final perceptions of intergroup empathic norms) to the content participants were actually exposed to (i.e., actual intergroup empathic norms in the empathy learning task), Experiment 7 was able to shed light on how learning biases – most probably, *negative priors*, *negativity bias* and *reciprocity bias* – might single-handedly contribute to the negativity and inaccuracy of intergroup empathic norm perceptions even when people are not exposed to unrepresentative intergroup content. Following this initial research, more work is needed to gain an understanding of *how* (and not only what) people learn about groups, group relations and group norms and to better map the learning biases and their potential to distort the relationship between learning inputs (i.e., the actual views, emotions, and behaviours people are exposed to) and learning outputs (i.e., the perceptions of views, emotions, and behaviours people form on the basis of their exposure).

Finally, while Experiment 7 sought to understand how *individuals'* learning is affected by biases and how *individuals* respond to varying intergroup dynamics, a productive avenue for future research concerns combining this study of individual-level psychological mechanisms with the study of group-level phenomena. In particular, future research should examine how large-scale group-level phenomena (e.g., the dynamics of conflict and benevolence) emerge when collections of individuals, fuelled with previously established individual-level psychological mechanisms (e.g., tendency to reciprocate outgroup members' behaviour towards their ingroup), interact (i.e., *agent-based modelling*, Chiang, 2021; Gray et al., 2014).

Broader theoretical considerations and future directions

This thesis has aimed to advance our understanding of the ways in which group category membership and induced or perceived group relations influence two related socio-affective processes: (a) judgments of facial expressions of emotion and (b) empathy. Previous sections of this chapter summarised the results of individual experiments, reviewed their theoretical and practical implications and offered recommendations for immediate follow-up research. In this final section, I will discuss broader theoretical considerations that have become salient over of the course of the research presented in this thesis and review outstanding questions which are derived from these considerations.

Intergroup empathy versus intergroup compassion

Given that the experiments reported in Chapters 3 and 4 mainly (but not exclusively[80]) focused on empathy, it is important to address recent criticisms of empathy in the literature. The main concern raised against empathy is that it "makes us [people] morally worse" (Bloom, 2017a, p. 30, see also Bloom 2017b), inclined to favour a single suffering individual over the suffering thousands (i.e., "innumeracy bias", Cameron et al., 2017) as well as members of our own groups over deserving, but socially distant others (i.e., "intergroup bias", Cikara et al., 2014). The criticisms of empathy do not stop there but continue to emphasise that empathy is aversive and can lead to burnout (Cameron et al., 2019; Hunt et al., 2017) and in extreme cases, motivate people to aggress against others (Buffone & Poulin, 2014). Under this view, empathy is seen as doing more harm than good, but what is the suggested alternative? Compassion – a more detached feeling of kindness and concern for others' welfare – has been suggested by the proponents of this view to serve as a

[80] Experiment 5 in Chapter 3 assessed participants' empathy and compassion alongside their perceptions of respective norms. The results were identical regardless of whether empathy or compassion were analysed.

more reliable guide for moral action (Bloom, 2017a). In light of these opinions, is intergroup empathy the right process interventions should be targeting or is intergroup compassion a better option?

In my opinion, favouring compassion over empathy (or vice versa) amplifies the differences between these processes at the expense of discussing their similarities, simplifies the relationship between them as well as the overall challenge of promoting general prosocial orientation across group lines. First, most of the points raised against empathy – its propensity to innumeracy and intergroup biases, as well as its aversiveness – have been shown to equally apply to compassion. Like empathy, compassion favours close over distant others (Bruneau et al., 2012; Scheffer et al., 2022) and breaks down as the number of suffering victims increases (Cameron, 2017). Furthermore, compassion is perceived as more taxing than empathy and people tend to avoid it as much as or even more than empathy (Scheffer et al., 2022).

Second, although recognised as distinct processes, empathy and compassion cannot be so easily decoupled in human experience as these criticisms suggest. For instance, in everyday life, people experience perspective taking, empathy and compassion in combination rather than in isolation, with empathy and compassion co-occurring in as much as 85% of the cases (Depow et al., 2021). Some accounts consider empathy to be a necessary precursor of compassion (Stevens & Taber, 2021) such that sharing an empathic target's suffering activates compassion which translates negative feelings to feelings of concern for the empathic target (Eisenberg et al., 1998; Engen & Singer, 2015). Therefore, the questions of how empathy and compassion are related and how these processes independently and jointly contribute to prosocial behaviour are active areas of inquiry in social psychology and neuroscience, rendering any preferences for one process over the other potentially premature.

Finally, these criticisms mostly point out potential problems with empathy "at baseline" while the research presented in this thesis is designed to identify ways in which some of these problems might be resolved. Even when applied to empathy-building interventions, these criticisms are most relevant to those interventions which explicitly instruct people to feel the pain of others (e.g., "Look at the person in the picture, and try to feel what this person feels. Share the suffering and experiences of this person.", Scheffer et al., 2022). Interventions adopting empathic norms, even when designed to primarily reflect empathy as valence matching between empathisers' and empathic targets' emotions (as in Experiment 7), are likely to communicate not just that empathy is normative, but also that kindness and general prosocial orientation towards others are normative (Nook et al., 2016). In other words, while empathy and compassion might be experientially (and conceptually) distinct, observers are likely to derive the same normative message from both public expressions of intergroup empathy and public expressions of intergroup compassion. To illustrate with an example, imagine going online and seeing that something negative has happened to a few outgroup members. Whether you then read online comments such as "I felt their pain as if it were my own", which more prototypically signals empathy, or such as "I care so much for these people, I hope they are safe soon", which more prototypically signals compassion, it is probable that you will infer the same normative message of care or prosociality. Therefore, the influence of intergroup empathic norms is likely to go beyond intergroup empathy to affect intergroup compassion and prosocial behaviour (Nook et al., 2016). In summary, recent criticisms of empathy and in favour of compassion (Bloom, 2017a; Bloom, 2017b) obscure the complex relationship between these processes and their general similarities. Particularly relevant for the current thesis, these criticisms do not apply to interventions adopting intergroup empathic norms as these interventions can be considered

to portray intergroup prosociality as normative, rather than just intergroup (affective) empathy.

Relational social categorisation

People are born into a world of pre-packaged group categories – gender categories, ethnicities, nationalities, professions – and learn about these categories both experientially (e.g., by playing with gender-specific toys, by observing residential segregation) and conceptually (e.g., by learning about a distant country without ever visiting it or interacting with its nationals). In addition to learning about group categories as such, people learn about relations within and between these categories – for example, growing up in Serbia, I quickly learnt that relations between ethnic Serbs and Greeks tend to be friendly (Armakolas & Triantafyllou, 2017) and relations between ethnic Serbs and Kosovar Albanians tend to be hostile (Maloku et al., 2016).

The current thesis largely mimicked such social environments with labelable group categories and varying group relations within and between them. While this approach was useful in that it examined relational features alongside defined (minimal or real-world) group categories, it should not be lost in translation that relations do not only come into play within and between recognised group categories, but also frequently form the very basis of social categorisation (Gershman & Cikara, 2020; Pietraszewski, 2022). In particular, people can detect groups by keeping track of relations between individuals. Using an example of open conflict, if one is a bystander observing an interaction where people A and B insult a person C after which a person D steps into person C's defence, one is likely to categorise people A and B into one group and people C and D into another (Pietraszewski, 2022). Using an example of emotional fit, if one expresses a strong negative emotion about an issue in the presence of other people after which people A and B express a similar emotion, while people C and D express a completely different emotion, one is likely to self-categorise with people

A and B over people C and D (Livingstone et al., 2016; see also Livingstone et al., 2011). In both examples, expressed emotions served as observable cues of the underlying relational configuration between people (Parkinson, 2021; Parkinson et al., 2005) with perception of "groups" ultimately arising not from explicit placement into some available group category[81] but from recognising how different individuals relate to oneself and each other.

To date, scholarship on the role of emotions in group processes and on the role of group processes in emotions, including the current thesis, has typically focused on one of the following questions: (a) how group category membership influences judgments of others' emotions and one's own emotional experience (e.g., Čehajić-Clancy et al., 2011; Smith & Mackie, 2015), and (b) what antecedents and consequences are associated with the experience of certain emotions in intergroup contexts (e.g., Halperin, 2014; Leonard et al., 2011; Maitner et al., 2007). In both lines of inquiry, emotions have predominantly been examined *alongside defined group categories*, while much less work has been devoted to the question of how emotions, with their relational implications, can serve as a basis for social categorisation to begin with (for a notable exception, see Livingstone et al., 2016; for an evolutionary account of how even racial group categories can stem from relational social categorisation, see Pietraszewski, 2020). Reiterating the point made earlier in this chapter, to fully realise the potential of this research direction, social psychology needs (a) a model of social relations in order to constrain a whole multitude of possible relational dimensions of interest, and (b) a better understanding of how emotions signal social relations and how social relations restrict emotions in order to map possible emotion-relation inferences (see Bryant et

[81] Critics may argue that group categories "attackers" and "defenders" could have become salient in the first example and group categories "negative opinion holders on issue X" and "positive opinion holders on issue X" could have become salient in the second example. In general, I do not oppose the idea that perceivers could have derived labels for inferred groups, but I emphasise that in these examples, labels should be seen as a product of *underlying relational computations*, not as a product of contextual activation of pre-existing group categories.

al., 2016; Martens et al., 2012). In summary, in addition to examining emotions in defined intra- and intergroup contexts, which is what the current thesis focused on, future research should further explore the question of how emotions allow one to carve such contexts in the first place and how relational configurations between people, signalled by emotions, interact with pre-packaged group categories (e.g., race, gender, nationality) to shape social categorisation.

Feature-oriented thinking about groups

All of the experiments reported in this thesis aimed to walk a fine line between (a) striving for their conclusions not to remain bounded by examined ethnic, political, or minimal group categories and (b) indiscriminately interpreting results in terms of "ingroups" and "outgroups" with little recognition that there might be meaningful contextual variability at play. For instance, Experiments 4 and 5 sought to provide insights into how *competitive* intergroup contexts shape perceptions of intergroup empathic norms, and not into how these perceptions operate among ethnic Croats and Serbs (Experiment 4), Leavers and Remainers (Experiment 5) or across all possible intergroup contexts regardless of the nature of intergroup relations. Similarly, even though Experiment 7 relied on the minimal group paradigm, which is notoriously difficult to generalise to real-world intergroup contexts, I went beyond minimal group categories ("Tigers", "Lions", "minimal ingroup", "minimal outgroup") to analyse *features* of groups (e.g., perceived deservingness of empathy), group relations (e.g., interdependence) and group environments (e.g., absence of punishment in the empathy learning task). Transparently discussing theoretically relevant contextual features should allow future researchers to expand on the results of Experiment 7 *regardless of which intergroup context they decide to study intergroup empathic norms in*.

Feature-oriented thinking about groups, which I tried to model in this thesis, is not unheard of in social cognition and social neuroscience research (see examples Cikara et al.,

2014; Cikara et al., 2022; illustrated in Chapter 1), and is even better represented in classical social psychological research inspired by the Social Identity Perspective (e.g., Giannakakis & Fritsche, 2011; Jackson et al., 1996; Mullen et al., 1992). However, at least in the fields of social cognition and social neuroscience, feature-oriented thinking about groups has remained rather secondary, mostly coming into use after an effect of interest has been documented among certain group categories and now needs to be interpreted (e.g., first we document that female facial expressions are evaluated as more happy than male facial expressions and then we take a step back to examine whether this tendency stems from differences in facial features or associated stereotypes). While group-category-centred evidence has been crucial in establishing the basic layout of intergroup phenomena and general social categorisation principles, I argue that psychological research has by now generated sufficient knowledge that we should be able to create space for deliberate theorising about contextual features in addition to group categories.

In summary, feature-oriented thinking about groups does not advocate completely abandoning the study of specific group categories and intergroup relations but argues that shifting the analytic lens from group categories (i.e., social identities as such) to features of groups, group relations and group environments should produce insights that might otherwise remain hidden behind the black-box of group category labels. Finally, although the current thesis illustrated feature-oriented thinking about groups by predominantly focusing on features of *group relations*, this should not create the false impression that this is the only class of features available to explore. Rather, depending on one's topic of interest, relevant contextual features could be anything from features of the physical environment to features of broader political and institutional discourses (Cikara et al., 2022).

Conclusion

Across seven experiments, this thesis has aimed to improve our understanding of how group category membership and induced or perceived group relations influence two closely related socio-affective processes: (a) judgments of facial expressions of emotion and (b) empathy. In doing so, this thesis presented preliminary evidence that minimal group biases in valence judgments of facial expressions cannot be accounted for by underlying coalitional relations and provided one of the most extensive examinations of intergroup empathic norms to date, outlining how these norms are subjectively perceived, learnt, and how they exert influence on one's own intergroup empathy and general intergroup feelings. These results shed light on the interplay between socio-affective and group processes and provide several methodological and conceptual recommendations for understanding and studying this interplay, particularly with regard to the role of social relations in helping people identify group structures and systematise the complexity of human interconnectedness. Finally, with this thesis, I join a growing number of researchers who have advocated moving beyond group categories as units of analysis and interpretation in social cognition and social neuroscience research and incorporating features of groups, group relations and group environments more meaningfully into empirical research and theorising.

REFERENCES

Abele, A. E., & Brack, S. (2013). Preference for other persons' traits is dependent on the kind of social relationship. *Social Psychology*, *44*, 84–94. https://doi.org/10.1027/1864-9335/a000138

Abele, A. E., & Wojciszke, B. (2014). Communal and Agentic Content in Social Cognition. *Advances in Experimental Social Psychology*, *50*, 195–255. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Adams, C. D., & Dickinson, A. (1981). Instrumental Responding following Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology*, *33*(2b), 109–121. https://doi.org/10.1080/14640748108400816

Aktar, E., Nimphy, C. A., Kret, M. E., Pérez-Edgar, K., Bögels, S. M., & Raijmakers, M. E. J. (2021). Pupil responses to dynamic negative facial expressions of emotion in infants and parents. *Developmental Psychobiology*, *63*(7), e22190. https://doi.org/10.1002/dev.22190

Allen, T. J., Sherman, J. W., Conrey, F. R., & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology*, *45*(5), 1081–1087. https://doi.org/10.1016/j.jesp.2009.06.002

Allidina, S., & Cunningham, W. A. (2021). Avoidance begets avoidance: A computational account of negative stereotype persistence. *Journal of Experimental Psychology: General*, *150*(10), 2078–2099. https://doi.org/10.1037/xge0001037

Amin, M. R., & Faghih, R. T. (2019). Tonic and Phasic Decomposition of Skin Conductance Data: A Generalized-Cross-Validation-Based Block Coordinate Descent Approach.

2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 745–749. https://doi.org/10.1109/EMBC.2019.8857074

Amodio, D. M. (2008). The social neuroscience of intergroup relations. *European Review of Social Psychology*, *19*(1), 1–54. https://doi.org/10.10463280801927937

Amodio, D. M., & Berg, J. J. (2022). The neural underpinnings of intergroup bias. In P. Molenberghs, *The Neuroscience of Intergroup Relations* (pp. 3–15). Routledge. https://doi.org/10.4324/9781003042426-1

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, *84*(4), 738–753. https://doi.org/10.1037/0022-3514.84.4.738

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human Brain Mapping*, *23*(4), 200–209. https://doi.org/10.1002/hbm.20048

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Arioli, M., Cattaneo, Z., Ricciardi, E., & Canessa, N. (2021). Overlapping and specific neural correlates for empathizing, affective mentalizing, and cognitive mentalizing: A coordinate-based meta-analytic study. *Human Brain Mapping*, *42*(14), 4777–4804. https://doi.org/10.1002/hbm.25570

Armakolas, I., & Triantafyllou, G. (2017). Greece and EU enlargement to the Western Balkans: Understanding an ambivalent relationship. *Southeast European and Black Sea Studies*, *17*(4), 611–629. https://doi.org/10.1080/14683857.2017.1407399

Asendorpf, J., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology, 74,* 1531–1544. https://doi.org/10.1037/0022-3514.74.6.1531

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality. *Psychological Bulletin, 130*, 80–114. https://doi.org/10.1037/0033-2909.130.1.80

Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*, *34*(12), 3168–3181. https://doi.org/10.1002/hbm.22133

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, *140*(6), 1556–1581. https://doi.org/10.1037/a0037737

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition*, 59–66. https://doi.org/10.1109/FG.2018.00019

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164–175. https://doi.org/10.1016/j.evolhumbehav.2013.02.002

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., ... & Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629-1643. https://doi.org/10.1177/0146167212457953

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models*. arXiv. https://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. arXiv. http://arxiv.org/abs/1406.5823

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*, 105–118. https://doi.org/10.1037/0022-3514.72.1.105

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of psychophysiology* (pp. 142–162). Cambridge University Press.

Beaupré, M. G., & Hess, U. (2006). An Ingroup Advantage for Confidence in Emotion Recognition Judgments: The Moderating Effect of Familiarity With the Expressions

of Outgroup Members. *Personality and Social Psychology Bulletin*, *32*(1), 16–26. https://doi.org/10.1177/0146167205277097

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Bell, L., Vogt, J., Willemse, C., Routledge, T., Butler, L. T., & Sakaki, M. (2018). Beyond Self-Report: A Review of Physiological and Neuroscientific Methods to Investigate Consumer Behavior. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01655

Benedek, M., & Kaernbach, C. (2010a). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, *47*(4), 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x

Benedek, M., & Kaernbach, C. (2010b). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Ben-Shakhar, G. (1985). Standardization Within Individuals: A Simple Method to Neutralize Individual Differences in Skin Conductance. *Psychophysiology*, *22*(3), 292–299. https://doi.org/10.1111/j.1469-8986.1985.tb01603.x

Bereczkei, T., Birkas, B., & Kerekes, Z. (2007). Public charity offer as a proximate factor of evolved reputation-building strategy: An experimental analysis of a real-life situation. *Evolution and Human Behavior*, *28*(4), 277–284. https://doi.org/10.1016/j.evolhumbehav.2007.04.002

Berman, J. Z., & Silver, I. (2022). Prosocial behavior and reputation: When does doing good lead to looking good? *Current Opinion in Psychology*, *43*, 102–107. https://doi.org/10.1016/j.copsyc.2021.06.021

Berndsen, M., & Tiggemann, M. (2020). Multiple versus single immoral acts: An immoral person evokes more schadenfreude than an immoral action. *Motivation and Emotion*, *44*(5), 738–754. https://doi.org/10.1007/s11031-020-09843-5

Berndsen, M., Tiggemann, M., & Chapman, S. (2017). "It wasn't your fault, but …...": Schadenfreude about an undeserved misfortune. *Motivation and Emotion*, *41*(6), 741–748. https://doi.org/10.1007/s11031-017-9639-1

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The Cross-Category Effect: Mere Social Categorization Is Sufficient to Elicit an Own-Group Bias in Face Recognition. *Psychological Science*, *18*(8), 706–712. https://doi.org/10.1111/j.1467-9280.2007.01964.x

Betancourt, H. (1990). An Attribution-Empathy Model of Helping Behavior: Behavioral Intentions and Judgments of Help-Giving. *Personality and Social Psychology Bulletin*, *16*(3), 573–591. https://doi.org/10.1177/0146167290163015

Bian, L., & Baillargeon, R. (2022). When Are Similar Individuals a Group? Early Reasoning About Similarity and In-Group Support. *Psychological Science*, *33*(5), 752–764. https://doi.org/10.1177/09567976211055185

Bică, A. (2022). Peer pressure: Conformity outweighs reciprocity in social anxiety. *Current Psychology*, 1–8. https://doi.org/10.1007/s12144-022-03021-1

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bieberstein, F. von, Essl, A., & Friedrich, K. (2021). Empathy: A clue for prosocialty and driver of indirect reciprocity. *PLOS ONE, 16*(8), e0255071. https://doi.org/10.1371/journal.pone.0255071

Bijlstra, G., Holland, R., Dotsch, R., & Wigboldus, D. (2018). Stereotypes and Prejudice Affect the Recognition of Emotional Body Postures. *Emotion*, *19*(2), 189–199. https://doi.org/10.1037/emo0000438

Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K., & Wigboldus, D. H. J. (2014). Stereotype Associations and Emotion Recognition. *Personality and Social Psychology Bulletin*, *40*(5), 567–577. https://doi.org/10.1177/0146167213520458

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, *46*(4), 657–663. https://doi.org/10.1016/j.jesp.2010.03.006

Bijlstra, G., Kleverwal, D., van Lent, T., & Holland, R. W. (2019). Evaluations versus stereotypes in emotion recognition: A replication and extension of Craig and Lipp's (2018) study on facial age cues. *Cognition and Emotion, 33*(2), 386–389. https://doi.org/10.1080/02699931.2018.1526778

Bjornsdottir, R. T., Cheso, D., & Rule, N. (2021). Beyond categories: Perceiving sexual attraction from faces. *British Journal of Psychology*, *113*(1), 226–247. https://doi.org/10.1111/bjop.12523

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes How We Look at Faces. *PLOS ONE*, *3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, *29*(1), 43–74.

https://doi.org/10.1002/(SICI)1099-0992(199902)29:1<43::AID-EJSP913>3.0.CO;2-Z

Bloom, P. (2017a). Empathy and its discontents. *Trends in Cognitive Sciences, 21*(1), 24-31. https://doi.org/10.1016/j.tics.2016.11.004

Bloom, P. (2017b). *Against empathy: The case for rational compassion*. Random House.

Bodenhausen G. V., Kang S. K., Peery D. (2012). Social categorization and the perception of social groups. In Fiske S. T., Macrae C. N. (Eds.), *Sage handbook of social cognition* (pp. 311–329). SAGE.

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, Configural, and Holistic Face-Processing Strategies Evoke Different Scan Patterns. *Perception, 38*(10), 1508–1521. https://doi.org/10.1068/p6117

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion Recognition: The Role of Featural and Configural Face Information. *Quarterly Journal of Experimental Psychology*, *66*(12), 2426–2442. https://doi.org/10.1080/17470218.2013.789065

Borinca, I., Andrighetto, L., Valsecchi, G., & Berent, J. (2021a). Ingroup norms shape understanding of outgroup prosocial behaviors. *Group Processes & Intergroup Relations*, 1368430220987604. https://doi.org/10.1177/1368430220987604

Borinca, I., Falomir-Pichastor, J. M., Andrighetto, L., & Durante, F. (2021b). Outgroup prejudice and perceptions of prosocial intergroup behaviors. *European Journal of Social Psychology, 51*(1), 40–53. https://doi.org/10.1002/ejsp.2712

Boucsein, W. (2012). *Electrodermal Activity*. Springer Science & Business Media.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment* (pp. 29–46). Oxford University Press.

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875), 20180116. https://doi.org/10.1098/rspb.2018.0116

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, *49*(1), 1017–1034.

Brambilla, M., & Riva, P. (2017). Predicting pleasure at others' misfortune: Morality trumps sociability and competence in driving deservingness and schadenfreude. *Motivation and Emotion*, *41*(2), 243–253. https://doi.org/10.1007/s11031-016-9594-2

Brewer, M. B. (1999). The psychology of prejudice: In-group love or out-group hate. *Journal of Social Issues*, 429–444.

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 215–232). Psychology Press.

Brkić, D., Ng-Cordell, E., O'Brien, S., Martin, J., Scerif, G., Astle, D., & Baker, K. (2022). FarmApp: A new assessment of cognitive control and memory for children and young people with neurodevelopmental difficulties. *Child Neuropsychology*, *28*(8), 1097–1115. https://doi.org/10.1080/09297049.2022.2054968

Brod, G., Greve, A., Jolles, D., Theobald, M., & Galeano-Keiner, E. M. (2022). Explicitly predicting outcomes enhances learning of expectancy-violating information. *Psychonomic Bulletin & Review*, *29*(6), 2192–2201. https://doi.org/10.3758/s13423-022-02124-x

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*, *2*(8), 581–591. https://doi.org/10.1038/s41562-018-0376-6

Brown, L. M., Bradley, M. M., & Lang, P. J. (2006). Affective reactions to pictures of ingroup and outgroup members. *Biological Psychology*, *71*(3), 303–311. https://doi.org/10.1016/j.biopsycho.2005.06.003

Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*, *5*(8), 998–1008. https://doi.org/10.1038/s41562-021-01066-z

Brubaker, R. (2006). *Ethnicity without groups*. Harvard University Press.

Bruckmüller, S., & Abele, A. E. (2013). The density of the Big Two: How are agency and communion structurally represented? *Social Psychology*, *44*, 63–74. https://doi.org/10.1027/1864-9335/a000145

Bruneau, E. G., Cikara, M., & Saxe, R. (2015). Minding the gap: Narrative descriptions about mental states attenuate parochial empathy. *PLOS ONE*, *10*(10), e0140838. https://doi.org/10.1371/journal.pone.0140838

Bruneau, E. G., Cikara, M., & Saxe, R. (2017). Parochial Empathy Predicts Reduced Altruism and the Endorsement of Passive Harm. *Social Psychological and Personality Science*, *8*(8), 934–942. https://doi.org/10.1177/1948550617693064

Bruneau, E. G., Dufour, N., & Saxe, R. (2012). Social cognition in members of conflict groups: Behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 717–730. https://doi.org/10.1098/rstb.2011.0293

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology*, *48*(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*, 123–152. https://doi.org/10.1037/h0043805

Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., Petersen, M. B., Bickham, S. T., Bolyanatz, A., Chavez, B., De Smet, D., Díaz, C., Fančovičová, J., Fux, M., Giraldo-Perez, P., Hu, A., Kamble, S. V., Kameda, T., Li, N. P., … Zhou, Y. (2016). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences*, *113*(17), 4682–4687. https://doi.org/10.1073/pnas.1524993113

Buffone, A. E. K., & Poulin, M. J. (2014). Empathy, Target Distress, and Neurohormone Genes Interact to Predict Aggression for Others–Even Without Provocation. *Personality and Social Psychology Bulletin*, *40*(11), 1406–1422. https://doi.org/10.1177/0146167214549320

Bruneau, E. G., Cikara, M., & Saxe, R. (2017). Parochial empathy predicts reduced altruism and the endorsement of passive harm. *Social Psychological and Personality Science*, *8*(8), 934–942. https://doi.org/10.1177/1948550617693064

Cameron, C. D. (2017). Compassion collapse: Why we are numb to numbers? In E. M. Seppälä, E. Simon-Thomas, S. L. Brown, M. C. Worline, C. D. Cameron, & J. R.

Doty (Eds.), *The Oxford Handbook of Compassion Science* (pp. 271–282). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190464684.013.20

Cameron, C. D., Conway, P., & Scheffer, J. A. (2022). Empathy regulation, prosociality, and moral judgment. *Current Opinion in Psychology*, *44*, 188–195. https://doi.org/10.1016/j.copsyc.2021.09.011

Cameron, C. D., Cunningham, W., Saunders, B., & Inzlicht, M. (2017). *The ends of empathy: Constructing empathy from value-based choice*. PsyArXiv. https://doi.org/10.31234/osf.io/d99bp

Cameron, C. D., Harris, L. T., & Payne, B. K. (2016). The Emotional Cost of Humanity: Anticipated Exhaustion Motivates Dehumanization of Stigmatized Targets. *Social Psychological and Personality Science*, *7*(2), 105–112. https://doi.org/10.1177/1948550615604453

Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, *148*, 962–976. https://doi.org/10.1037/xge0000595

Cameron, C. D., Scheffer, J. A., Hadjiandreou, E., & Anderson, S. (2022). Motivated empathic choices. *Advances in Experimental Social Psychology*, *66*, 191–279. https://doi.org/10.1016/bs.aesp.2022.04.005

Candini, M., Battaglia, S., Benassi, M., di Pellegrino, G., & Frassinetti, F. (2021). The physiological correlates of interpersonal space. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-82223-2

Carr, J. W., Pescuma, V. N., Furlan, M., Ktori, M., & Crepaldi, D. (2022). Algorithms for the automated correction of vertical drift in eye-tracking data. *Behavior Research Methods, 54*(1), 287–310. https://doi.org/10.3758/s13428-021-01554-0

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics, 11*(3), e1460. https://doi.org/10.1002/wics.1460

Cavazza, N., Pagliaro, S., & Guidetti, M. (2014). Antecedents of Concern for Personal Reputation: The Role of Group Entitativity and Fear of Social Exclusion. *Basic and Applied Social Psychology, 36*(4), 365–376. https://doi.org/10.1080/01973533.2014.925453

Čehajić-Clancy, S., Effron, D. A., Halperin, E., Liberman, V., & Ross, L. D. (2011). Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. *Journal of Personality and Social Psychology, 101*(2), 256–270. https://doi.org/10.1037/a0023936

Chang, L. W., Krosch, A. R., & Cikara, M. (2016). Effects of intergroup threat on mind, brain, and behavior. *Current Opinion in Psychology, 11*, 69–73. https://doi.org/10.1016/j.copsyc.2016.06.004

Cheon, B. K., Im, D., Harada, T., Kim, J.-S., Mathur, V. A., Scimeca, J. M., Parrish, T. B., Park, H. W., & Chiao, J. Y. (2011). Cultural influences on neural basis of intergroup empathy. *NeuroImage, 57*(2), 642–650. https://doi.org/10.1016/j.neuroimage.2011.04.031

Chiang, Y. S. (2021). Indirect Reciprocity for Mitigating Intergroup Hostility: A Vignette Experiment and an Agent-based Model on Intergroup Relations between Mainland Chinese and Taiwanese. *Journal of Conflict Resolution*, *65*, 403–426. https://doi.org/10.1177/0022002720962209

Chiu, C. Y., Hong, Y. Y., & Dweck, C. (1997). Lay Dispositionism and Implicit Theories of Personality. *Journal of Personality and Social Psychology*, *73*, 19–30. https://doi.org/10.1037/0022-3514.73.1.19

Chiroro, P., & Valentine, T. (1995). An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*(4), 105–109. https://doi.org/10.1111/1467-8721.01242

Ciarleglio, A. (2021). Measures of variability and precision in statistics: Appreciating, untangling and applying concepts. *BJPsych Advances*, *27*(2), 137–139. https://doi.org/10.1192/bja.2020.41

Cikara, M. (2015). Intergroup Schadenfreude: Motivating participation in collective violence. *Current Opinion in Behavioral Sciences*, *3*, 12–17. https://doi.org/10.1016/j.cobeha.2014.12.007

Cikara, M. (2021). Causes and consequences of coalitional cognition. *Advances in Experimental Social Psychology*, *64*, 65–128. https://doi.org/10.1016/bs.aesp.2021.04.002

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and Them: Intergroup Failures of Empathy. *Current Directions in Psychological Science*, *20*(3), 149–153. https://doi.org/10.1177/0963721411408713

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of Experimental Social Psychology*, *55*, 110–125. https://doi.org/10.1016/j.jesp.2014.06.007

Cikara, M., & Fiske, S. T. (2012). Stereotypes and Schadenfreude: Affective and Physiological Markers of Pleasure at Outgroup Misfortunes. *Social Psychological and Personality Science*, *3*(1), 63–71. https://doi.org/10.1177/1948550611409245

Cikara, M., Martinez, J. E., & Lewis, N. A. (2022). Moving beyond social categories by incorporating context in social psychological theory. *Nature Reviews Psychology*, *1*(9), 537–549. https://doi.org/10.1038/s44159-022-00079-3

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, *9*(3), 245-274. https://doi.org/10.1177/1745691614527464

Clark, B. M., Siddle, D. A. T., & Bond, N. W. (1992). Effects of social anxiety and facial expression on habituation of the electrodermal orienting response. *Biological Psychology*, *33*(2), 211–223. https://doi.org/10.1016/0301-0511(92)90033-Q

Coll, M. P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience and Biobehavioral Reviews*, *83*, 132–139. https://doi.org/10.1016/j.neubiorev.2017.10.009

Contreras-Huerta, L. S., Baker, K. S., Reynolds, K. J., Batalha, L., & Cunnington, R. (2013). Racial Bias in Neural Empathic Responses to Pain. *PLOS ONE, 8*(12), e84001. https://doi.org/10.1371/journal.pone.0084001

Converse, P. E., Dotson, J. D., Hoag, W. J., McGee III, W. H., & McGee, I. I. I. (1980). *American social attitudes data sourcebook, 1947-1978*. Harvard University Press.

Conway, J. R., Coll, M., Cuve, H. C., Koletsi, S., Bronitt, N., Catmur, C., & Bird, G. (2020). Understanding how minds vary relates to skill in inferring mental states, personality, and intelligence. *Journal of Experimental Psychology: General, 149*(6), 1032–1047. https://doi.org/10.1037/xge0000704

Cooley, E., & Payne, B. K. (2019). A group is more than the average of its parts: Why existing stereotypes are applied more to the same individuals when viewed in groups than when viewed alone. *Group Processes & Intergroup Relations, 22*(5), 673–687. https://doi.org/10.1177/1368430218756494

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*, 75–93. https://doi.org/10.1037/emo0000302

Correll, J., & Hudson, S. M. (2020). An error in the analysis of "An eye for the I". *Journal of Personality and Social Psychology, 119*, 1030–1036. https://doi.org/10.1037/pspa0000200

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Correll, J., Wittenbrink, B., Park, B., Judd, C. M., & Goyle, A. (2011). Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of Experimental Social Psychology*, *47*(1), 184–189. https://doi.org/10.1016/j.jesp.2010.08.017

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research*, *146*(1–2), 167–174. https://doi.org/10.1016/j.bbr.2003.09.025

Craig, B. M., & Lipp, O. V. (2018a). Facial age cues and emotional expression interact asymmetrically: Age cues moderate emotion categorisation. *Cognition and Emotion*, *32*(2), 350–362. https://doi.org/10.1080/02699931.2017.1310087

Craig, B. M., & Lipp, O. V. (2018b). The influence of multiple social categories on emotion perception. *Journal of Experimental Social Psychology*, *75*, 27–35. https://doi.org/10.1016/j.jesp.2017.11.002

Craig, M. A., & Lee, M. M. (2022). Status-based coalitions: Hispanic growth affects Whites' perceptions of political support from Asian Americans. *Group Processes & Intergroup Relations*, *25*(3), 661–681. https://doi.org/10.1177/13684302211032286

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. https://doi.org/10.1037/0022-3514.82.3.359

Croatian Bureau of Statistics. (2005). *Naselja i stanovništvo Republike Hrvatske od 1857. do 2001. godine* [Settlements and population of the Republic of Croatia from 1857 to 2001]. Retrived March 29, 2022, from https://www.dzs.hr/hrv/dbhomepages/naselja%20i%20stanovnistvo%20republike%20hrvatske/naselja%20i%20stanovnistvo%20republike%20hrvatske.htm

Croatian Bureau of Statistics. (2011). *Population by ethnicity - Detailed classification, 2011 census*. Retrieved March 22, 2022, from https://www.dzs.hr/eng/censuses/census2011/results/htm/e01_01_05/E01_01_05.html

Crockett, M., Özdemir, Y., & Fehr, E. (2014). The Value of Vengeance and the Demand for Deterrence. *Journal of Experimental Psychology: General*, *143*(6), 2279–2286. https://doi.org/10.1037/xge0000018

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), 1–33. https://doi.org/10.1348/014466608X314935

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Cushman, F., & Gershman, S. (2019). Editors' Introduction: Computational Approaches to Social Cognition. *Topics in Cognitive Science*, *11*(2), 281–298. https://doi.org/10.1111/tops.12424

Cutler, J., Wittmann, M. K., Abdurahman, A., Hargitai, L. D., Drew, D., Husain, M., & Lockwood, P. L. (2021). Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-24576-w

Cuve, H. C., Gao, Y., & Fuse, A. (2018). Is it avoidance or hypoarousal? A systematic review of emotion recognition, eye-tracking, and psychophysiological studies in

young adults with autism spectrum conditions. *Research in Autism Spectrum Disorders, 55*, 1–13. https://doi.org/10.1016/j.rasd.2018.07.002

Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2021). Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods, 54*(2), 1027–1049. https://doi.org/10.3758/s13428-021-01654-x

Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., & Cottrell, G. W. (2010). Evidence and a Computational Explanation of Cultural Differences in Facial Expression Recognition. *Emotion, 10*(6), 874–893. https://doi.org/10.1037/a0020019

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. https://doi.org/10.1037/0022-3514.44.1.113

Davis, M. H., & Begovic, E. (2014). Empathy-related interventions. In A. C. Parks & S. M. Schueller (Eds). The Wiley Blackwell handbook of positive psychological interventions (pp. 111–134). https://doi.org/10.1002/9781118315927.ch6

Davis, J. R., & Gold, G. J. (2011). An examination of emotional empathy, attributions of stability, and the link between perceived remorse and forgiveness. *Personality and Individual Differences*, *50*(3), 392–397. https://doi.org/10.1016/j.paid.2010.10.031

Dawson, M. E., Filion, D. L., & Schell, A. M. (1989). Is Elicitation of the Autonomic Orienting Response Associated With Allocation of Processing Resources? *Psychophysiology*, *26*(5), 560–572. https://doi.org/10.1111/j.1469-8986.1989.tb00710.x

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science, 4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119

Decety, J., Echols, S., & Correll, J. (2010). The Blame Game: The Effect of Responsibility and Social Stigma on Empathy for Pain. *Journal of Cognitive Neuroscience*, *22*(5), 985–997. https://doi.org/10.1162/jocn.2009.21266

Decker, J. H., Lourenco, F. S., Doll, B. B., & Hartley, C. A. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 310–320. https://doi.org/10.3758/s13415-014-0332-5

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, *18*(7), 337–339. https://doi.org/10.1016/j.tics.2014.04.008

De Leersnyder, J. (2017). Emotional acculturation: A first review. *Current Opinion in Psychology*, *17*, 67–73. https://doi.org/10.1016/j.copsyc.2017.06.007

Depow, G. J., Francis, Z., & Inzlicht, M. (2021). The Experience of Empathy in Everyday Life. *Psychological Science*, *32*(8), 1198–1213. https://doi.org/10.1177/0956797621995202

Devine, P. G. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. https://doi.org/10.1016/j.tics.2006.08.008

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Diehl, M. (1988). Social identity and minimal groups: The effects of interpersonal and intergroup attitudinal similarity on intergroup discrimination. *British Journal of*

Social Psychology, *27*(4), 289–300. https://doi.org/10.1111/j.2044-8309.1988.tb00833.x

Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. *European Review of Social Psychology*, *1*(1), 263–292. https://doi.org/10.1080/14792779108401864

Dindo, L., & Fowles, D. C. (2008). The skin conductance orienting response to semantic stimuli: Significance can be independent of arousal. *Psychophysiology*, *45*(1), 111–118. https://doi.org/10.1111/j.1469-8986.2007.00604.x

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Dotsch, R. O. N., Wigboldus, D. H. J., & Knippenberg, A. V. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, *43*(1), 116–125. https://doi.org/10.1002/ejsp.1928

Drwecki, B. B., Moore, C. F., Ward, S. E., & Prkachin, K. M. (2011). Reducing racial disparities in pain treatment: The role of empathy and perspective-taking. *Pain*, *152*(5), 1001–1006. https://doi.org/10.1016/j.pain.2010.12.005

Dunham, Y. (2011). An angry = Outgroup effect. *Journal of Experimental Social Psychology*, *47*(3), 668–671. https://doi.org/10.1016/j.jesp.2011.01.003

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780–793. https://doi.org/10.1016/j.tics.2018.06.004

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "Minimal" Group Affiliations in Children. *Child Development*, *82*(3), 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Durik, A. M., Hyde, J. S., Marks, A. C., Roy, A. L., Anaya, D., & Schultz, G. (2006). Ethnicity and Gender Stereotypes of Emotion. *Sex Roles, 54*(7), 429–445. https://doi.org/10.1007/s11199-006-9020-4

Durrheim, K., Quayle, M., Tredoux, C. G., Titlestad, K., & Tooke, L. (2016). Investigating the Evolution of Ingroup Favoritism Using a Minimal Group Interaction Paradigm: The Effects of Inter- and Intragroup Interdependence. *PLOS ONE*, *11*(11), e0165974. https://doi.org/10.1371/journal.pone.0165974

Dweck, C. S., Chiu, C., & Hong, Y. (1995a). Implicit Theories and Their Role in Judgments and Reactions: A Word From Two Perspectives. *Psychological Inquiry*, *6*(4), 267–285. https://doi.org/10.1207/s15327965pli0604_1

Dweck, C. S., Chiu, C., & Hong, Y. (1995b). Implicit Theories: Elaboration and Extension of the Model. *Psychological Inquiry*, *6*(4), 322–333. https://doi.org/10.1207/s15327965pli0604_12

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, *24*, 96–102. https://doi.org/10.1016/j.lindif.2012.12.020

Edwards, K., & Smith, E. E. (1996). A Disconfirmation Bias in the Evaluation of Arguments. *Journal of Personality and Social Psychology, 71*(1), 5–24.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. https://doi.org/10.1037/a0022758

Eisenberg, N., Wentzel, M., & Harris, J. D. (1998). The Role of Emotionality and Regulation in Empathy-Related Responding. *School Psychology Review*, *27*(4), 506–521. https://doi.org/10.1080/02796015.1998.12085934

Ekman, P., & Friesen W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System: Investigators guide*. Salt Lake City, UT: Research Nexus.

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate facial movement. *Child Development*, *51*(3), 886–891. 10.2307/1129478

Elfenbein, H. A. (2013). Nonverbal Dialects and Accents in Facial Expressions of Emotion. *Emotion Review*, *5*(1), 90–96. https://doi.org/10.1177/1754073912451332

Elfenbein, H. A. (2015). In-Group Advantage and Other-Group Bias in Facial Emotion Recognition. In M. K. Mandal & A. Awasthi (Eds.), *Understanding Facial Expressions in Communication* (pp. 57–71). Springer India. https://doi.org/10.1007/978-81-322-1934-7_4

Elfenbein, H. A., & Ambady, N. (2002). Is there an in-group advantage in emotion recognition? *Psychological Bulletin*, *128*(2), 243–249. https://doi.org/10.1037/0033-2909.128.2.243

Elfenbein, H. A., & Ambady, N. (2003). Universals and Cultural Differences in Recognizing Emotions. *Current Directions in Psychological Science*, *12*(5), 159–164. https://doi.org/10.1111/1467-8721.01252

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity.

European Journal of Social Psychology, *29*, 371–389. https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/33.3.CO;2-L

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-Year Olds, but Not Chimpanzees, Attempt to Manage Their Reputations. *PLOS ONE*, *7*(10), e48433. https://doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, *20*, 92–95. https://doi.org/10.1016/j.copsyc.2017.08.024

Engen, H. G., & Singer, T. (2015). Compassion-based emotion regulation up-regulates experienced positive affect and associated neural networks. *Social Cognitive and Affective Neuroscience*, *10*(9), 1291–1301. https://doi.org/10.1093/scan/nsv008

Enock, F. E., Flavell, J. C., Tipper, S. P., & Over, H. (2021). No convincing evidence outgroups are denied uniquely human characteristics: Distinguishing intergroup preference from trait-based dehumanization. *Cognition*, *212*, 104682. https://doi.org/10.1016/j.cognition.2021.104682

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, *9*, 15–25. https://www.frontiersin.org/articles/10.3389/fnbeh.2015.00015

Ewbank, M. P., Jennings, C., & Calder, A. J. (2009). Why are you angry with me? Facial expressions of threat influence perception of gaze direction. *Journal of Vision, 9*(12), 16–27. https://doi.org/10.1167/9.12.16

Fannes, G., & Claeys, A.-S. (2022). Putting empathic feelings into words during times of crisis: The impact of differential verbal empathy expressions on organizational reputation. *Public Relations Review, 48*(2), 102183. https://doi.org/10.1016/j.pubrev.2022.102183

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

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15*(2), 115–141. https://doi.org/10.1080/02699930125908

Ferguson, N., & McKeown, S. (2016). Social identity theory and intergroup conflict in Northern Ireland. In S. McKeown, R. Haji, & N. Ferguson (Eds.), *Understanding Peace and Conflict Through Social Identity Theory: Contemporary Global Perspectives* (pp. 215-227). Springer, Cham.

Feather, N. T. (2006). Deservingness and emotions: Applying the structural model of deservingness to the analysis of affective reactions to outcomes. *European Review of Social Psychology, 17*(1), 38–73. https://doi.org/10.1080/10463280600662321

Feather, N. T., & Mckee, I. R. (2014). Deservingness, liking relations, schadenfreude, and other discrete emotions in the context of the outcomes of plagiarism. *Australian Journal of Psychology, 66*(1), 18–27. https://doi.org/10.1111/ajpy.12030

Feather, N. T., Wenzel, M., & McKee, I. R. (2013). Integrating multiple perspectives on schadenfreude: The role of deservingness and emotions. *Motivation and Emotion*, *37*(3), 574–585. https://doi.org/10.1007/s11031-012-9331-4

FeldmanHall, O., & Nassar, M. R. (2021). The computational challenge of social learning. *Trends in Cognitive Sciences*, *25*(12), 1045–1057. https://doi.org/10.1016/j.tics.2021.09.002

Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, *43*, 1–6. https://doi.org/10.1016/j.copsyc.2021.06.005

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision-making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research* (pp. 163–184). Psychology Press.

Filion, D. L., Dawson, M. E., Schell, A. M., & Hazlett, E. A. (1991). The Relationship Between Skin Conductance Orienting and the Allocation of Processing Resources. *Psychophysiology*, *28*(4), 410–424. https://doi.org/10.1111/j.1469-8986.1991.tb00725.x

Fiske, A. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*, 689–723. https://doi.org/10.1037//0033-295X.99.4.689

Flippen, A. R., Hornstein, H. A., Siegal, W. E., & Weitzman, E. A. (1996). A Comparison of Similarity and Interdependence as Triggers for In-Group Formation. *Personality and Social Psychology Bulletin*, *22*(9), 882–893. https://doi.org/10.1177/0146167296229003

Folz, J., Fiacchino, D., Nikolić, M., van Steenbergen, H., & Kret, M. E. (2022). Reading Your Emotions in My Physiology? Reliable Emotion Interpretations in Absence of a Robust Physiological Resonance. *Affective Science*, *3*(2), 480–497. https://doi.org/10.1007/s42761-021-00083-5

Fowler, Z., Law, K. F., & Gaesser, B. (2021). Against empathy bias: The moral value of equitable empathy. *Psychological Science*, *32*(5), 766–779. https://doi.org/10.1177/0956797620979965

Frahm, O. (2012). Defining the Nation: National Identity in South Sudanese Media Discourse. *Africa Spectrum*, *47*(1), 21–49. https://doi.org/10.1177/000203971204700102

Freeman, J. B. (2018). Doing Psychological Science by Hand. *Current Directions in Psychological Science*, *27*(5), 315–323. https://doi.org/10.1177/0963721417746793

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Freeman, J. B., & Johnson, K. L. (2016). More Than Meets the Eye: Split-Second Social Perception. *Trends in Cognitive Sciences*, *20*(5), 362–374. https://doi.org/10.1016/j.tics.2016.03.003

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The Neural Basis of Contextual Influences on Face Categorization. *Cerebral Cortex*, *25*(2), 415–422. https://doi.org/10.1093/cercor/bht238

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology*, *61*, 237–287. https://doi.org/10.1016/bs.aesp.2019.09.005

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. A. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*, *24*, 83–91. https://doi.org/10.1016/j.copsyc.2018.10.003

Friesen, J. P., Kawakami, K., Vingilis-Jaremko, L., Caprara, R., Sidhu, D. M., Williams, A., Hugenberg, K., Rodríguez-Bailón, R., Cañadas, E., & Niedenthal, P. (2019). Perceiving happiness in an intergroup context: The role of race and attention to the eyes in differentiating between true and false smiles. *Journal of Personality and Social Psychology*, *116*(3), 375–395. https://doi.org/10.1037/pspa0000139

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frucht, R. C. (Ed.). (2005). *Eastern Europe: An introduction to the people, lands, and culture*. ABC-CLIO.

Gable, S. L., Reis, H. T., Impett, E. A., & Asher, E. R. (2004). What Do You Do When Things Go Right? The Intrapersonal and Interpersonal Benefits of Sharing Positive Events. *Journal of Personality and Social Psychology*, *87*(2), 228–245. https://doi.org/10.1037/0022-3514.87.2.228

Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: Categorization, reciprocation, or fear? *Journal of Personality and Social Psychology*, *79*, 77–94. https://doi.org/10.1037/0022-3514.79.1.77

Gaertner, L., & Schopler, J. (1998). Perceived ingroup entitativity and intergroup bias: An interconnection of self and others. *European Journal of Social Psychology*, *28*(6), 963–980. https://doi.org/10.1002/(SICI)1099-0992(1998110)28:6<963::AID-EJSP905>3.0.CO;2-S

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics: Theory, Research, and Practice*, *4*, 98–114. https://doi.org/10.1037/1089-2699.4.1.98

Gaertner, S. L., Dovidio, J. F., & Houlette, M. A. (2010). Social categorization. In J. F. Dovidio, P. Glick, & M. Hewstone (Eds.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 526–543).

Gaesser, B., Shimura, Y., & Cikara, M. (2020). Episodic simulation reduces intergroup bias in prosocial intentions and behavior. *Journal of Personality and Social Psychology*, *118*(4), 683–705. https://doi.org/10.1037/pspi0000194

Gamond, L., & Cattaneo, Z. (2016). The dorsomedial prefrontal cortex plays a causal role in mediating in-group advantage in emotion recognition: A TMS study. *Neuropsychologia*, *93*, 312–317. https://doi.org/10.1016/j.neuropsychologia.2016.11.011

Gamond, L., Vilarem, E., Safra, L., Conty, L., & Grèzes, J. (2017). Minimal group membership biases early neural processing of emotional expressions. *European Journal of Neuroscience*, *46*(10), 2584–2595. https://doi.org/10.1111/ejn.13735

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behavior Research Methods*, *52*(5), 2232–2255. https://doi.org/10.3758/s13428-020-01374-8

Gershman, S. J., & Cikara, M. (2020). Social-Structure Learning. *Current Directions in Psychological Science*, *29*(5), 460–466. https://doi.org/10.1177/0963721420924481

Giannakakis, A. E., & Fritsche, I. (2011). Social identities, group norms, and threat: On the malleability of ingroup bias. *Personality and Social Psychology Bulletin*, *37*(1), 82–93. 10.1177/0146167210386120

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523–536. https://doi.org/10.3758/s13415-015-0347-6

Gillespie, A., Howarth, C. S., & Cornish, F. (2012). Four problems for researchers using social categories. *Culture & Psychology*, *18*(3), 391–402. https://doi.org/10.1177/1354067X12446236

Gitter, A. G., Black, H., & Mostofsky, D. (1972). Race and Sex in the Perception of Emotion. *Journal of Social Issues*, *28*(4), 63–78. https://doi.org/10.1111/j.1540-4560.1972.tb00046.x

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. https://doi.org/10.1016/j.cognition.2014.11.040

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212–216. 10.1257/000282806777211658

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in Cross-Race Face Learning: Insights From Eye Movements and Pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1105–1122. https://doi.org/10.1037/a0016548

Gosselin, P., Perron, M., & Beaupré, M. (2010). The voluntary control of facial action units in adults. *Emotion*, *10*(2), 266–271. https://doi.org/10.1037/a0017748

Gray, K., Rand, D. G., Ert, E., Lewis, K., Hershman, S., & Norton, M. I. (2014). The Emergence of "Us and Them" in 80 Lines of Code: Modeling Group Genesis in Homogeneous Populations. *Psychological Science*, *25*(4), 982–990. https://doi.org/10.1177/0956797614521816

Greco, M., & Baenninger, R. (1991). Effects of yawning and related activities on skin conductance and heart rate. *Physiology & Behavior*, *50*(5), 1067–1069. https://doi.org/10.1016/0031-9384(91)90439-U

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Green, S. R., Kragel, P. A., Fecteau, M. E., & LaBar, K. S. (2014). Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis. *International Journal of Psychophysiology*, *91*(3), 186–193. https://doi.org/10.1016/j.ijpsycho.2013.10.015

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41. https://doi.org/10.1037/a0015575

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

Gubler, J. R., Halperin, E., & Hirschberger, G. (2015). Humanizing the outgroup in contexts of protracted intergroup conflict. *Journal of Experimental Political Science*, *2*(1), 36–46. https://doi.org/10.1017/xps.2014.20

Gutsell, J. N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, *46*(5), 841-845. https://doi.org/10.1016/j.jesp.2010.03.011

Gutsell, J. N., & Inzlicht, M. (2012). Intergroup differences in the sharing of emotive states: Neural evidence of an empathy gap. *Social Cognitive and Affective Neuroscience*, *7*(5), 596–603. https://doi.org/10.1093/scan/nsr035

Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Current Opinion in Psychology*, *24*, 92–97. https://doi.org/10.1016/j.copsyc.2018.09.001

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233–1235. https://doi.org/10.1038/nn.4080

Hackel, L. M., Kogon, D., Amodio, D. M., & Wood, W. (2022). Group value learned through interactions with members: A reinforcement learning account. *Journal of Experimental Social Psychology*, *99*, 104267. https://doi.org/10.1016/j.jesp.2021.104267

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23. https://doi.org/10.1016/j.jesp.2013.12.001

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, *65*(6), 1105–1118. https://doi.org/10.1037/0022-3514.65.6.1105

Haimovitz, K., & Dweck, C. S. (2016). Parents' Views of Failure Predict Children's Fixed and Growth Intelligence Mind-Sets. *Psychological Science*, *27*(6), 859–869. https://doi.org/10.1177/0956797616639727

Halberstadt, A. G., Cooke, A. N., Garner, P. W., Hughes, S. A., Oertwig, D., & Neupert, S. D. (2020). Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion*, *22*(3), 403–417. https://doi.org/10.1037/emo0000756

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, *159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442

Hall, J. A., Schwartz, R., & Duong, F. (2021). How do laypeople define empathy? *The Journal of Social Psychology*, *161*(1), 5–24. https://doi.org/10.1080/00224545.2020.1796567

Halperin, E. (2014). Emotion, Emotion Regulation, and Conflict Resolution. *Emotion Review*, *6*(1), 68–76. https://doi.org/10.1177/1754073913491844

Hargreaves Heap, S. P., & Zizzo, D. J. (2009). The Value of Groups. *American Economic Review*, *99*(1), 295–323. https://doi.org/10.1257/aer.99.1.295

Haslam, N., & Fiske, A. P. (1999). Relational models theory: A confirmatory factor analysis. *Personal Relationships*, *6*(2), 241–250. https://doi.org/10.1111/j.1475-6811.1999.tb00190.x

Haslam, N., & Fiske, A. P. (2004). Social expertise: Theory of mind or theory of relationships?. In N. Haslam (Ed.), *Relational Models Theory* (pp. 163-180). Psychology Press.

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, 5(1), 60–65. 10.1177/1754073912451331

Hasson, Y., Amir, E., Sobol-Sarag, D., Tamir, M., & Halperin, E. (2022). Using performance art to promote intergroup prosociality by cultivating the belief that empathy is unlimited. *Nature Communications*, *13*(1), 1–15. https://doi.org/10.1038/s41467-022-35235-z

Hasson, Y., Tamir, M., Brahms, K. S., Cohrs, J. C., & Halperin, E. (2018). Are Liberals and Conservatives Equally Motivated to Feel Empathy Toward Others? *Personality and Social Psychology Bulletin*, *44*(10), 1449–1459. https://doi.org/10.1177/0146167218769867

Hawkins, R. X., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in Cognitive Sciences*, *23*(2), 158–169. https://doi.org/10.1016/j.tics.2018.11.003

Hein, G., Engelmann, J. B., Vollberg, M. C., & Tobler, P. N. (2016). How learning shapes the empathic brain. *Proceedings of the National Academy of Sciences*, *113*(1), 80–85. https://doi.org/10.1073/pnas.1514539112

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural Responses to Ingroup and Outgroup Members' Suffering Predict Individual Differences in Costly Helping. *Neuron*, *68*(1), 149–160. https://doi.org/10.1016/j.neuron.2010.09.003

Hepach, R., Vaish, A., Grossmann, T., & Tomasello, M. (2016). Young Children Want to See Others Get the Help They Need. *Child Development*, *87*(6), 1703–1714. https://doi.org/10.1111/cdev.12633

Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology*, *115*(1), 16–29. https://doi.org/10.1016/j.jecp.2012.12.013

Hertz, U. (2021). Learning how to behave: cognitive learning processes account for asymmetries in adaptation to social norms. *Proceedings of the Royal Society B*, *288*(1952), 20210293. https://doi.org/10.1098/rspb.2021.0293

Hess. U. (2011). Peripheral psychophysiological methods. In K. C. Klauer, C. Stahl, & A. Voss (Eds.), *Cognitive Methods in Social Psychology* (pp. 265- 303). Guilford Press.

Hess, U., Adams, R. B., & Kleck, R. E. (2009). The face is not an empty canvas: How facial expressions interact with facial appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3497–3504. https://doi.org/10.1098/rstb.2009.0165

Hess, E. H., & Polt, J. M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. *Science*, *132*(3423), 349–350. https://doi.org/10.1126/science.132.3423.349

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hess, E. H., Seltzer, A. L., & Shlien, J. M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology*, *70*, 165–168. https://doi.org/10.1037/h0021978

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, *27*(5), 856–881. https://doi.org/10.3758/s13423-020-01715-w

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, *48*(4), 1694–1712. https://doi.org/10.3758/s13428-015-0676-y

Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. R. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology, 100*(3), 492–506. https://doi.org/10.1037/a0021562

Hobolt, S. B., Leeper, T. J., & Tilley, J. (2021). Divided by the vote: Affective polarization in the wake of the Brexit referendum. *British Journal of Political Science, 51*(4), 1476-1493. https://doi.org/10.1017/S0007123420000125

Hodges, S. D., Kiel, K. J., Kramer, A. D. I., Veach, D., & Villanueva, B. R. (2010). Giving Birth to Empathy: The Effects of Similar Experience on Empathic Accuracy, Empathic Concern, and Perceived Empathy. *Personality and Social Psychology Bulletin*, *36*(3), 398–409. https://doi.org/10.1177/0146167209350326

Hogg, M. A. (2016). Social Identity Theory. In S. McKeown, R. Haji, & N. Ferguson (Eds.), *Understanding Peace and Conflict Through Social Identity Theory: Contemporary Global Perspectives* (pp. 3–17). Springer International Publishing. https://doi.org/10.1007/978-3-319-29869-6_1

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The Social Identity Perspective: Intergroup Relations, Self-Conception, and Small Groups. *Small Group Research*, *35*(3), 246–276. https://doi.org/10.1177/1046496404263424

Hogg, M. A., & Smith, J. R. (2007). Attitudes in social context: A social identity perspective. *European Review of Social Psychology*, *18*(1), 89–131. https://doi.org/10.1080/10463280701592070

Hogg, M. A., Terry, D. J., & White, K. M. (1995). A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly*, *58*(4), 255–269. https://doi.org/10.2307/2787127

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52. https://doi.org/10.1145/2168556.2168563

Hong, Y., & Ratner, K. G. (2020). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology, 120*(3), 576–600. https://doi.org/10.1037/pspa0000255

Hong, Y. Y., Chiu, C. Y., Dweck, C., Lin, D., & Wan, W. (1999). Implicit Theories, Attributions, and Coping: A Meaning System Approach. *Journal of Personality and Social Psychology, 77*, 588–599. https://doi.org/10.1037/0022-3514.77.3.588

Hoogland, C. E., Ryan Schurtz, D., Cooper, C. M., Combs, D. J. Y., Brown, E. G., & Smith, R. H. (2015). The joy of pain and the pain of joy: In-group identification predicts schadenfreude and gluckschmerz following rival groups' fortunes. *Motivation and Emotion, 39*(2), 260–281. https://doi.org/10.1007/s11031-014-9447-9

Houwer, J. D. (2009). How do People Evaluate Objects? A Brief Review. *Social and Personality Psychology Compass, 3*(1), 36–48. https://doi.org/10.1111/j.1751-9004.2008.00162.x

Hsiao, J. H., & Cottrell, G. (2008). Two Fixations Suffice in Face Recognition. *Psychological Science, 19*(10), 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x

Hudson, S. T. J. (2022). *Empathic and counter-empathic emotions shape social hierarchy based on group positionality*. PsyArXiv. https://doi.org/10.31234/osf.io/w3x8y

Huddy, L. (2002). Context and meaning in social identity theory: A response to Oakes. *Political Psychology, 23*(4), 825–838. https://www.jstor.org/stable/3792369

Huey, E. B. (1908). *The psychology and pedagogy of reading*. Macmillan.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing Prejudice: Implicit Prejudice and the Perception of Facial Threat. *Psychological Science*, *14*(6), 640–643. https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x

Hugenberg, K., & Corneille, O. (2009). Holistic Processing Is Tuned for In-Group Faces. *Cognitive Science*, *33*(6), 1173–1181. https://doi.org/10.1111/j.1551-6709.2009.01048.x

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340. https://doi.org/10.1016/j.jesp.2006.02.010

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*, 1168–1187. https://doi.org/10.1037/a0020463

Hunt, P. A., Denieffe, S., & Gooney, M. (2017). Burnout and its relationship to empathy in nursing: A review of the literature. *Journal of Research in Nursing*, *22*(1–2), 7–22. https://doi.org/10.1177/1744987116678902

Hutcherson, C., Seppala, E., & Gross, J. (2008). Loving-Kindness Meditation Increases Social Connectedness. *Emotion*, *8*, 720–724. https://doi.org/10.1037/a0013237

Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E., & Galinsky, A. D. (2018). The critical role of second-order normative beliefs in predicting energy conservation. *Nature Human Behaviour*, *2*(10), 757–764. https://doi.org/10.1038/s41562-018-0434-0

Jackson, J. W. (2002). Intergroup Attitudes as a Function of Different Dimensions of Group Identification and Perceived Intergroup Conflict. *Self and Identity*, *1*(1), 11–33. https://doi.org/10.1080/152988602317232777

Jackson, L. A., Sullivan, L. A., Harnish, R., & Hodge, C. N. (1996). Achieving positive social identity: Social mobility, social creativity, and permeability of group boundaries. *Journal of Personality and Social Psychology*, *70*(2), 241–254. https://doi.org/10.1037/0022-3514.70.2.241

Jessen, S., Altvater-Mackensen, N., & Grossmann, T. (2016). Pupillary responses reveal infants' discrimination of facial emotions independent of conscious perception. *Cognition*, *150*, 163–169. https://doi.org/10.1016/j.cognition.2016.02.010

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct? *Emotion*, *16*(8), 1107–1116. https://doi.org/10.1037/emo0000228

Jospe, K., Genzer, S., klein Selle, N., Ong, D., Zaki, J., & Perry, A. (2020). The contribution of linguistic and visual cues to physiological synchrony and empathic accuracy. *Cortex*, *132*, 296–308. https://doi.org/10.1016/j.cortex.2020.09.001

Juarez, M. C., Ishii, L., Nellis, J. C., Bater, K., Huynh, P. P., Fung, N., Darrach, H., Russell, J. O., & Ishii, M. (2019). Objectively measuring social attention of thyroid neck scars and transoral surgery using eye tracking. *The Laryngoscope*, *129*(12), 2789–2794. https://doi.org/10.1002/lary.27933

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kamans, E., Gordijn, E. H., Oldenhuis, H., & Otten, S. (2009). What I think you see is what you get: Influence of prejudice on assimilation to negative meta-stereotypes among Dutch Moroccan teenagers. *European Journal of Social Psychology*, *39*(5), 842–851. https://doi.org/10.1002/ejsp.593

Kang, S.-M., & Lau, A. S. (2013). Revisiting the Out-Group Advantage in Emotion Recognition in a Multicultural Society: Further Evidence for the In-Group Advantage. *Emotion*, *13*(2), 203–215. https://doi.org/10.1037/a0030013

Kanske, P., Böckler, A., & Singer, T. (2015). Models, Mechanisms and Moderators Dissociating Empathy and Theory of Mind. *Current Topics in Behavioral Neurosciences*, *30,* 193–206. https://doi.org/10.1007/7854_2015_412

Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, *11*(9), 1383–1392. https://doi.org/10.1093/scan/nsw052

Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Kawakami, K., Friesen, J. P., Williams, A., Vingilis-Jaremko, L., Sidhu, D. M., Rodriguez-Bailón, R., Cañadas, E., & Hugenberg, K. (2021). Impact of perceived interpersonal similarity on attention to the eyes of same-race and other-race faces. *Cognitive*

Research: Principles and Implications, *6*(1), 1–16. https://doi.org/10.1186/s41235-021-00336-8

Kawakami, K., Williams, A., Sidhu, D., Choma, B. L., Rodriguez-Bailón, R., Cañadas, E., Chung, D., & Hugenberg, K. (2014). An eye for the I: Preferential attention to the eyes of ingroup members. *Journal of Personality and Social Psychology*, *107*, 1–20. https://doi.org/10.1037/a0036838

Kawamura, Y., & Kusumi, T. (2020). Altruism does not always lead to a good reputation: A normative explanation. *Journal of Experimental Social Psychology*, *90*, 104021. https://doi.org/10.1016/j.jesp.2020.104021

Keysers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, *18*(4), 163–166. https://doi.org/10.1016/j.tics.2013.12.011

Kinner, V. L., Kuchinke, L., Dierolf, A. M., Merz, C. J., Otto, T., & Wolf, O. T. (2017). What our eyes tell us about feelings: Tracking pupillary responses during emotion regulation processes. *Psychophysiology*, *54*(4), 508–518. https://doi.org/10.1111/psyp.12816

Kleider-Offutt, H. M., Bond, A. D., Williams, S. E., & Bohil, C. J. (2018). When a face type is perceived as threatening: Using general recognition theory to understand biased categorization of Afrocentric faces. *Memory & Cognition*, *46*(5), 716–728. https://doi.org/10.3758/s13421-018-0801-0

Klimecki, O. M. (2019). The Role of Empathy and Compassion in Conflict Resolution. *Emotion Review*, *11*(4), 310–325. https://doi.org/10.1177/1754073919838609

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive and Affective Neuroscience*, *9*(6), 873–879. https://doi.org/10.1093/scan/nst060

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068–1078. https://doi.org/10.1111/ejn.12859

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*, 675–709. https://doi.org/10.1037/pspa0000046

Kommattam, P., Jonas, K., & Fischer, A. (2017). We Are Sorry, They Don't Care: Misinterpretation of Facial Embarrassment Displays in Arab-White Intergroup Contexts. *Emotion*, *17*(4), 658–668. https://doi.org/10.1037/emo0000261

Kommattam, P., Jonas, K. J., & Fischer, A. H. (2019). Perceived to feel less: Intensity bias in interethnic emotion perception. *Journal of Experimental Social Psychology*, *84*, 103809. https://doi.org/10.1016/j.jesp.2019.04.007

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., Fink, G. R., & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604. https://doi.org/10.1007/s00429-014-0803-z

Kranton, R., Pease, M., Sanders, S., & Huettel, S. (2020). Deconstructing bias in social preferences reveals groupy and not-groupy behavior. *Proceedings of the National Academy of Sciences*, *117*(35), 21185–21193. https://doi.org/10.1073/pnas.1918952117

Krautheim, J. T., Dannlowski, U., Steines, M., Neziroğlu, G., Acosta, H., Sommer, J., Straube, B., & Kircher, T. (2019). Intergroup empathy: Enhanced neural resonance for ingroup facial emotion in a shared neural production-perception network. *NeuroImage*, *194*, 182–190. https://doi.org/10.1016/j.neuroimage.2019.03.048

Kredentser, M. S., Fabrigar, L. R., Smith, S. M., & Fulton, K. (2012). Following What People Think We Should Do Versus What People Actually Do: Elaboration as a Moderator of the Impact of Descriptive and Injunctive Norms. *Social Psychological and Personality Science*, *3*(3), 341–347. https://doi.org/10.1177/1948550611420481

Kret, M., Roelofs, K., Stekelenburg, J., & de Gelder, B. (2013). Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience, 7*, 810–829. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00810

Krüger, M., Bartels, W., & Krist, H. (2020). Illuminating the Dark Ages: Pupil Dilation as a Measure of Expectancy Violation Across the Life Span. *Child Development*, *91*(6), 2221–2236. https://doi.org/10.1111/cdev.13354

Kteily, N., Hodson, G., & Bruneau, E. (2016a). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, *110*(3), 343–370. https://doi.org/10.1037/pspa0000044

Kteily, N., Hodson, G., & Bruneau, E. (2016b). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, *110*(3), 343–370. https://doi.org/10.1037/pspa0000044

Ku, G., Malhotra, D., & Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes*, *96*(2), 89–103. https://doi.org/10.1016/j.obhdp.2004.10.001

Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences*, *118*(6), e2008389118. https://doi.org/10.1073/pnas.2008389118

Kuhn, M., Gerlicher, A. M. V., & Lonsdorf, T. B. (2022). Navigating the manyverse of skin conductance response quantification approaches – A direct comparison of trough-to-peak, baseline correction, and model-based approaches in Ledalab and PsPM. *Psychophysiology*, *59*(9), e14058. https://doi.org/10.1111/psyp.14058

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 53(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*(4), 522–544. https://doi.org/10.1037/0033-2909.129.4.522

Kunstman, J. W., Tuscherer, T., Trawalter, S., & Lloyd, E. P. (2016). What Lies Beneath? Minority Group Members' Suspicion of Whites' Egalitarian Motivation Predicts Responses to Whites' Smiles. *Personality and Social Psychology Bulletin*, *42*(9), 1193–1205. https://doi.org/10.1177/0146167216652860

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*(4), 917–940. https://doi.org/10.1037/a0030811

Kuzmanovic, B., & Rigoux, L. (2017). Valence-Dependent Belief Updating: Computational Validation. *Frontiers in Psychology, 8,* 1087–1097. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01087

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science, 4*(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Lam, T. C. M., Kolomitro, K., & Alamparambil, F. C. (2011). Empathy Training: Methods, Evaluation Practices, and Validity. *Journal of Multidisciplinary Evaluation, 7*(16), 162–200.

Landry, A. P., Ihm, E., & Schooler, J. W. (2022). Hated but still human: Metadehumanization leads to greater hostility than metaprejudice. *Group Processes & Intergroup Relations, 25*(2), 315–334. https://doi.org/10.1177/1368430220979035

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lavin, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience, 7,* 218–229. https://www.frontiersin.org/articles/10.3389/fnbeh.2013.00218

Lazerus, T., Ingbretsen, Z. A., Stolier, R. M., Freeman, J. B., & Cikara, M. (2016). Positivity bias in judging ingroup members' emotional expressions. *Emotion*, *16*(8), 1117–1125. https://doi.org/10.1037/emo0000227

Leach, C. W., Van Zomeren, M., Zebel, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., ... & Spears, R. (2008). Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Lee, S.L., Chiu, C.-Y., & Chan, T.-K. (2005). Some boundary conditions of the expressor culture effect in emotion recognition: Evidence from Hong Kong Chinese perceivers. *Asian Journal of Social Psychology*, *8*(3), 224–243. https://doi.org/10.1111/j.1467-839X.2005.00169.x

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279–286. https://doi.org/10.1038/s41562-019-0766-4

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1822), 20200143. https://doi.org/10.1098/rstb.2020.0143

Lehmann, K., Böckler, A., Klimecki, O., Müller-Liebmann, C., & Kanske, P. (2022). Empathy and correct mental state inferences both promote prosociality. *Scientific Reports*, *12*(1), 1–8. https://doi.org/10.1038/s41598-022-20855-8

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*, 1–33. https://www.jstatsoft.org/article/view/v069i01

Leonard, D. J., Mackie, D. M., & Smith, E. R. (2011). Emotional responses to intergroup apology mediate intergroup forgiveness and retribution. *Journal of Experimental Social Psychology*, *47*(6), 1198–1206. https://doi.org/10.1016/j.jesp.2011.05.002

Leonardelli, G. J., & Toh, S. M. (2015). Social Categorization in Intergroup Contexts: Three Kinds of Self-Categorization. *Social and Personality Psychology Compass*, *9*(2), 69–87. https://doi.org/10.1111/spc3.12150

Leong, Y. C., & Zaki, J. (2018). Unrealistic optimism in advice taking: A computational account. *Journal of Experimental Psychology: General*, *147*(2), 170–189. https://doi.org/10.1037/xge0000382

Leuchs, L., Schneider, M., & Spoormaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, *56*(1), e13283. https://doi.org/10.1111/psyp.13283

Levendusky, M. S., & Malhotra, N. (2016). (Mis)perceptions of Partisan Polarization in the American Public. *Public Opinion Quarterly*, *80*, 378–391. https://doi.org/10.1093/poq/nfv045

Levin, D. (2001). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition. *Journal of Experimental Psychology. General*, *129*, 559–574. https://doi.org/10.1037/0096-3445.129.4.559

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.

Li, X., McAllister, D. J., Ilies, R., & Gloor, J. L. (2019). Schadenfreude: A Counternormative Observer Response to Workplace Mistreatment. *Academy of Management Review*, *44*(2), 360–376. https://doi.org/10.5465/amr.2016.0134

Liberman, Z., Howard, L. H., Vasquez, N. M., & Woodward, A. L. (2018). Children's expectations about conventional and moral behaviors of ingroup and outgroup members. *Journal of Experimental Child Psychology*, *165*, 7–18. https://doi.org/10.1016/j.jecp.2017.03.003

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, *21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2017). The Role of a 'Common Is Moral' Heuristic in the Stability and Change of Moral Norms. *Journal of Experimental Psychology: General*, *147*(2), 228–242. https://doi.org/10.1037/xge0000365

Lindström, B., Selbing, I., Molapour, T., & Olsson, A. (2014). Racial Bias Shapes Social Reinforcement Learning. *Psychological Science*, *25*(3), 711–719. https://doi.org/10.1177/0956797613514093

Livingstone, A. G., Shepherd, L., Spears, R., & Manstead, A. S. R. (2016). "Fury, us": Anger as a basis for new group self-categories. *Cognition and Emotion*, *30*(1), 183–192. https://doi.org/10.1080/02699931.2015.1023702

Livingstone, A. G., Spears, R., Manstead, A. S. R., Bruder, M., & Shepherd, L. (2011). We feel, therefore we are: Emotion as a basis for self-categorization and social action. *Emotion*, *11*(4), 754–767. https://doi.org/10.1037/a0023223

Livingstone, A. G., Windeatt, S., Nesbitt, L., Kerry, J., Barr, S. A., Ashman, L., Ayers, R., Bibby, H., Boswell, E., Brown, J., Chiu, M., Cowie, E., Doherr, E., Douglas, H., Durber, L., Ferguson, M., Ferreira, M., Fisk, I., Fleming, B., … Wu, J.-C. (2020). Do you get us? A multi-experiment, meta-analytic test of the effect of felt understanding

in intergroup relations. *Journal of Experimental Social Psychology*, *91*, 104028. https://doi.org/10.1016/j.jesp.2020.104028

Lloyd, E. P., Lloyd, A. R., McConnell, A. R., & Hugenberg, K. (2022). Race Deficits in Pain Detection: Medical Providers and Laypeople Fail to Accurately Perceive Pain Authenticity Among Black People. *Social Psychological and Personality Science*, *13*(5), 895–905. https://doi.org/10.1177/19485506211045887

Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology*, *39*, 773–783. https://doi.org/10.1037/0022-3514.39.5.773

Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour—A reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761–771. https://doi.org/10.1093/scan/nsaa040

Long, E. L., Cuve, H. C., Conway, J. R., Catmur, C., & Bird, G. (2022). Novel theory of mind task demonstrates representation of minds in mental state inference. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-25490-x

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lucas, B. J., & Kteily, N. S. (2018). (Anti-)egalitarianism differentially predicts empathy for members of advantaged versus disadvantaged groups. *Journal of Personality and Social Psychology, 114*, 665–692. https://doi.org/10.1037/pspa0000112

Lueke, A., & Gibson, B. (2015). Mindfulness Meditation Reduces Implicit Age and Race Bias: The Role of Reduced Automaticity of Responding. *Social Psychological and Personality Science*, *6*(3), 284–291. https://doi.org/10.1177/1948550614559651

Lykstad, J., Reddy, V., & Hanna, A. (2022). Neuroanatomy, Pupillary Dilation Pathway. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK535421/

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Ma, Q., Meng, L., & Shen, Q. (2016). Deliberate or unintended: Intentions modulate empathic responses to others' economic payoffs in social interactions. *Social Neuroscience*, 1–12. https://doi.org/10.1080/17470919.2016.1223746

Maitner, A. T., Mackie, D. M., & Smith, E. R. (2007). Antecedents and Consequences of Satisfaction and Guilt Following Ingroup Aggression. *Group Processes & Intergroup Relations*, *10*(2), 223–237. https://doi.org/10.1177/1368430207075154

Maloku, E., Derks, B., Laar, C. V., & Ellemers, N. (2016). Building national identity in newborn Kosovo: Challenges of integrating national identity with ethnic identity among Kosovar Albanians and Kosovar Serbs. In S. McKeown, R. Haji, & N. Ferguson (Eds.), *Understanding Peace and Conflict Through Social Identity Theory: Contemporary Global Perspectives* (pp. 245-260). Springer, Cham.

Manrique, H. M., Zeidler, H., Roberts, G., Barclay, P., Walker, M., Samu, F., Fariña, A., Bshary, R., & Raihani, N. (2021). The psychological foundations of reputation-based

cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1838), 20200287. https://doi.org/10.1098/rstb.2020.0287

Marsh, A. A. (2022). Getting our affect together: Shared representations as the core of empathy. *Emotion Review*, 17540739221107029.

Martens, J. P., Tracy, J. L., & Shariff, A. F. (2012). Status signals: Adaptive benefits of displaying and observing the nonverbal expressions of pride and shame. *Cognition and Emotion*, *26*(3), 390–406. https://doi.org/10.1080/02699931.2011.645281

Masten, C. L., Gillen-O'Neel, C., & Brown, C. S. (2010). Children's intergroup empathic processing: The roles of novel ingroup identification, situational distress, and social anxiety. *Journal of Experimental Child Psychology*, *106*, 115–128. https://doi.org/10.1016/j.jecp.2010.01.002

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Automatic facial responses to briefly presented emotional stimuli in autism spectrum disorder. *Biological Psychology*, *94*(2), 397–407. https://doi.org/10.1016/j.biopsycho.2013.08.004

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 16–32. https://doi.org/10.5334/joc.18

Mathôt, S., Dalmaijer, E., Grainger, J., & Van der Stigchel, S. (2014). The pupillary light response reflects exogenous attention and inhibition of return. *Journal of Vision*, *14*(14), 7–17 . https://doi.org/10.1167/14.14.7

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary Responses to Words That Convey a Sense of Brightness or Darkness. *Psychological Science*, *28*(8), 1116–1124. https://doi.org/10.1177/0956797617702699

Mathôt, S., Siebold, A., Donk, M., & Vitu, F. (2015). Large pupils predict goal-driven eye movements. *Journal of Experimental Psychology: General*, *144*(3), 513–521. https://doi.org/10.1037/a0039168

Mathôt, S., & Vilotijević, A. (2022). Methods in Cognitive Pupillometry: Design, Preprocessing, and Statistical Analysis. *bioRxiv*.

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, *51*(4), 1468–1475. https://doi.org/10.1016/j.neuroimage.2010.03.025

Matsumoto, D., Olide, A., & Willingham, B. (2009). Is There an Ingroup Advantage in Recognizing Spontaneously Expressed Emotions? *Journal of Nonverbal Behavior*, *33*(3), 181–191. https://doi.org/10.1007/s10919-009-0068-z

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

Mazar, A., & Wood, W. (2018). Defining habit in psychology. In B. Verplanken (Ed.), *The Psychology of Habit* (pp. 13-29). Springer, Cham.

Mazar, A., & Wood, W. (2022). Illusory Feelings, Elusive Habits: People Overlook Habits in Explanations of Behavior. *Psychological Science*, *33*(4), 563–578. https://doi.org/10.1177/09567976211045345

McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150073. https://doi.org/10.1098/rstb.2015.0073

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147–151. https://doi.org/10.1016/j.cobeha.2015.04.006

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For Members Only: Ingroup Punishment of Fairness Norm Violations in the Ultimatum Game. *Social Psychological and Personality Science*, *5*(6), 662–670. https://doi.org/10.1177/1948550614527115

Menton, J. (2021, February 11). 'Looking down their nose at you': GameStop frenzy showed a fresh contempt for hedge funds. Why do Americans hate them?. USA Today. bit.ly/3BhqmWK

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proceedings of the National Academy of Sciences*, *119*(16), e2116851119. https://doi.org/10.1073/pnas.2116851119

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Michel, C., Corneille, O., & Rossion, B. (2007). Race Categorization Modulates Holistic Face Encoding. *Cognitive Science*, *31*(5), 911–924. https://doi.org/10.1080/03640210701530805

Miller, E., Krumhuber, E., & Dawel, A. (2020). Observers perceive the Duchenne marker as signaling only intensity for sad expressions, not genuine emotion. *Emotion*, *22*(5), 907–919. https://doi.org/10.1037/emo0000772

Misch, A., Paulus, M., & Dunham, Y. (2021). Anticipation of future cooperation eliminates minimal ingroup bias in children and adults. *Journal of Experimental Psychology: General*, *150*, 2036–2056. https://doi.org/10.1037/xge0000899

Molenberghs, P., Bosworth, R., Nott, Z., Louis, W. R., Smith, J. R., Amiot, C. E., Vohs, K. D., & Decety, J. (2014). The influence of group membership and individual differences in psychopathy and perspective taking on neural responses when punishing and rewarding others. *Human Brain Mapping*, *35*(10), 4989–4999. https://doi.org/10.1002/hbm.22527

Molenberghs, P., & Louis, W. R. (2018). Insights from fMRI studies into ingroup bias. *Frontiers in Psychology*, *9*, 1868–1878. https://doi.org/10.3389/fpsyg.2018.01868

Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-17286-2

Montalan, B., Lelard, T., Godefroy, O., & Mouras, H. (2012). Behavioral investigation of the influence of social categorization on empathy for pain: A minimal group paradigm study. *Frontiers in Psychology*, *3*, 1–5. https://doi.org/10.3389/fpsyg.2012.00389

Montrey, M., & Shultz, T. R. (2022). Copy the In-group: Group Membership Trumps Perceived Reliability, Warmth, and Competence in a Social-Learning Task. *Psychological Science*, *33*(1), 165–174. https://doi.org/10.1177/09567976211032224

Moore, L. J., Wilson, M. R., Vine, S. J., Coussens, A. H., & Freeman, P. (2013). Champ or chump? Challenge and threat states during pressurized competition. *Journal of Sport & Exercise Psychology*, *35*(6), 551–562. https://doi.org/10.1123/jsep.35.6.551

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*, *117*(26), 14864–14872. https://doi.org/10.1073/pnas.2001263117

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/10.1002/ejsp.2420220202

Müller-Pinzler, L., Czekalla, N., Mayer, A. V., Stolz, D. S., Gazzola, V., Keysers, C., Paulus, F. M., & Krach, S. (2019). Negativity-bias in forming beliefs about own abilities. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-50821-w

Murray, T., O'Brien, J., Sagiv, N., & Garrido, L. (2021). The role of stimulus-based cues and conceptual information in processing facial expressions of emotion. *Cortex*, *144*, 109–132. https://doi.org/10.1016/j.cortex.2021.08.007

Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D., & Willer, J.-C. (2005). Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, *43*(9), 1318–1328. https://doi.org/10.1016/j.neuropsychologia.2004.11.024

Nadler, A., & Liviatan, I. (2006). Intergroup reconciliation: Effects of adversary's expressions of empathy, responsibility, and recipients' trust. *Personality and Social Psychology Bulletin*, *32*(4), 459-470. https://doi.org/10.1177/0146167205276431

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nava, E., Romano, D., Grassi, M., & Turati, C. (2016). Skin conductance reveals the early development of the unconscious processing of emotions. *Cortex*, *84*, 124–131. https://doi.org/10.1016/j.cortex.2016.07.011

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, *4*(2), 199–203. https://doi.org/10.1002/wics.199

Ng, A. H., Steele, J. R., & Sasaki, J. Y. (2016). Will you remember me? Cultural differences in own-group face recognition biases. *Journal of Experimental Social Psychology*, *64*, 21–26. https://doi.org/10.1016/j.jesp.2016.01.003

Nook, E. C., Ong, D. C., Morelli, S. A., Mitchell, J. P., & Zaki, J. (2016). Prosocial conformity: Prosocial norms generalize across behavior and empathy. *Personality and Social Psychology Bulletin*, *42*(8), 1045–1062. https://doi.org/10.1177/0146167216649932

Nord, C. (1991). Scopos, loyalty, and translational conventions. *Target. International Journal of Translation Studies*, *3*(1), 91–109. https://doi.org/10.1075/target.3.1.06nor

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30*(2), 125–144. https://doi.org/10.1111/j.2044-8309.1991.tb00930.x

ODIHR (2021). *2021 Hate Crime Data*. https://hatecrime.osce.org/hate-crime-data?year=2021

Office of the High Commissioner for Human Rights. (2022). *Ukraine: civilian casualty update 22 August 2022*. https://www.ohchr.org/en/news/2022/08/ukraine-civilian-casualty-update-22-august-2022

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10*(9), 1095–1102. https://doi.org/10.1038/nn1968

Ong, D. C., Goodman, N. D., & Zaki, J. (2018). Happier than thou? A self-enhancement bias in emotion attribution. *Emotion, 18*(1), 116–126. https://doi.org/10.1037/emo0000309

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology, 11*, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010

Otten, S., & Epstude, K. (2006). Overlapping Mental Representations of Self, Ingroup, and Outgroup: Unraveling Self-Stereotyping and Self-Anchoring. *Personality and Social Psychology Bulletin*, *32*(7), 957–969. https://doi.org/10.1177/0146167206287254

Over, H., & McCall, C. (2018). Becoming us and them: Social learning and intergroup bias. *Social and Personality Psychology Compass*, *12*(4), e12384. https://doi.org/10.1111/spc3.12384

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*(3), 574–587. https://doi.org/10.1037/a0011989

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (2003). The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. *Journal of*

Psychosomatic Research, *55*(3), 269–275. https://doi.org/10.1016/S0022-3999(02)00578-0

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love." *Social Justice Research*, *26*(1), 81–96. https://doi.org/10.1007/s11211-012-0175-6

Parkinson, B. (2021). Heart to Heart: A Relation-Alignment Approach to Emotion's Social Effects. *Emotion Review*, *13*(2), 78–89. https://doi.org/10.1177/1754073921990733

Parkinson, B., Fischer, A. H., & Manstead, A. S. (2005). *Emotion in social relations: Cultural, group, and interpersonal processes*. Psychology Press.

Parks, C. D., & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology*, *99*, 303–310. https://doi.org/10.1037/a0018403

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1), 185–198. https://doi.org/10.1016/S1071-5819(03)00017-X

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. http://dx.doi.org/10.1037/met0000126

Penić, S., Dukes, D., Elcheroth, G., Jayakody, S., & Sander, D. (2021). Beyond personal empathy: Perceiving inclusive Empathy as socially shared predicts support for transitional justice mechanisms. *Affective Science*, *2*(4), 402–413. https://doi.org/10.1007/s42761-021-00086-2

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, *97*(1), 30–37. https://doi.org/10.1016/j.ijpsycho.2015.04.019

Pietraszewski, D. (2013). What is group psychology? Adaptations for mapping shared intentional stances. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 253–257). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199890712.003.0046

Pietraszewski,D. (2020). Intergroup processes: Principles from an evolutionary perspective. In P. Van Lange, E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 373 - 391). Guilford Press.

Pietraszewski, D. (2022). Toward a computational theory of social groups: A finite set of cognitive primitives for representing any and all social groups in the context of conflict. *Behavioral and Brain Sciences*, *45*, e97. https://doi.org/10.1017/S0140525X21000583

Pinter, B., & Greenwald, A. G. (2011). A comparison of minimal group induction procedures. *Group Processes & Intergroup Relations*, *14*(1), 81–98. https://doi.org/10.1177/1368430210375251

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 16–26. https://doi.org/10.1167/13.3.16

Pivetti, M., Camodeca, M., & Rapino, M. (2016). Shame, Guilt, and Anger: Their Cognitive, Physiological, and Behavioral Correlates. *Current Psychology*, *35*(4), 690–699. https://doi.org/10.1007/s12144-015-9339-5

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The Gender Stereotyping of Emotions. *Psychology of Women Quarterly*, *24*(1), 81–92. https://doi.org/10.1111/j.1471-6402.2000.tb01024.x

Platow, M. J., Grace, D. M., & Smithson, M. J. (2012). Examining the Preconditions for Psychological Group Membership: Perceived Social Interdependence as the Outcome of Self-Categorization. *Social Psychological and Personality Science*, *3*(1), 5–13. https://doi.org/10.1177/1948550611407081

Plötner, M., Over, H., Carpenter, M., & Tomasello, M. (2015). The effects of collaboration and minimal-group membership on children's prosocial behavior, liking, affiliation, and trust. *Journal of Experimental Child Psychology*, *139*, 161–173. https://doi.org/10.1016/j.jecp.2015.05.008

Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology*, *69*(3), 408–419. https://doi.org/10.1037/0022-3514.69.3.408

Porat, R., Halperin, E., & Tamir, M. (2016). What we want is what we get: Group-based emotional preferences and conflict resolution. *Journal of Personality and Social Psychology*, *110*(2), 167–190. https://doi.org/10.1037/pspa0000043

Powell, L. J. (2022). Adopted Utility Calculus: Origins of a Concept of Social Affiliation. *Perspectives on Psychological Science*, *17*(5), 1215–1233. https://doi.org/10.1177/17456916211048487

Powell, C. A. J., & Smith, R. H. (2013). Schadenfreude Caused by the Exposure of Hypocrisy in Others. *Self and Identity*, *12*(4), 413–431. https://doi.org/10.1080/15298868.2012.687013

Preckel, K., Kanske, P., & Singer, T. (2018). On the interaction of social affect and cognition: Empathy, compassion and theory of mind. *Current Opinion in Behavioral Sciences, 19*, 1–6. https://doi.org/10.1016/j.cobeha.2017.07.010

Presidents apologise over Croatian war (2003, September 10). *BBC News*. http://news.bbc.co.uk/2/hi/europe/3095774.stm

Press Association. (2022, March 7). Ukraine aid appeal tops £100m in UK. *The Guardian*. https://www.theguardian.com/world/2022/mar/07/ukraine-aid-appeal-tops-100m-in-uk

Preuschoff, K., 't Hart, B., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience, 5*, 115–125. https://www.frontiersin.org/articles/10.3389/fnins.2011.00115

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, *11*(1), 37–43. https://doi.org/10.1016/j.tics.2006.11.001

Proulx, T., Sleegers, W., & Tritt, S. M. (2017). The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology*, *70*, 69–79. https://doi.org/10.1016/j.jesp.2016.12.003

Prunty, J. E., Keemink, J. R., & Kelly, D. J. (2022). Infants show pupil dilatory responses to happy and angry facial expressions. *Developmental Science*, *25*(2), e13182. https://doi.org/10.1111/desc.13182

Psaltis, C., & Cakal, H. (2016). Social identity in a divided Cyprus. In S. McKeown, R. Haji, & N. Ferguson (Eds.), *Understanding Peace and Conflict Through Social Identity Theory: Contemporary Global Perspectives* (pp. 229-244). Springer, Cham.

Puurtinen, M., Heap, S., & Mappes, T. (2015). The joint emergence of group competition and within-group cooperation. *Evolution and Human Behavior*, *36*(3), 211–217. https://doi.org/10.1016/j.evolhumbehav.2014.11.005

Rabbie, J. M., & Horwitz, M. (1988). Categories versus groups as explanatory concepts in intergroup relations. *European Journal of Social Psychology*, *18*(2), 117–123. https://doi.org/10.1002/ejsp.2420180204

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology*, *19*(3), 171–202. https://doi.org/10.1002/ejsp.2420190302

Rahal, R.-M., & Fiedler, S. (2019). Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology*, *85*, 103842. https://doi.org/10.1016/j.jesp.2019.103842

Raihani, N. J., & Power, E. A. (2021). No good deed goes unpunished: The social costs of prosocial behaviour. *Evolutionary Human Sciences*, *3*, e40. https://doi.org/10.1017/ehs.2021.35

Randall, A. K., & Schoebi, D. (2015). Lean on me: Susceptibility to partner affect attenuates psychological distress over a 12-month period. *Emotion*, *15*, 201–210. https://doi.org/10.1037/emo0000043

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*(4), 450–472. https://doi.org/10.1016/j.cogdev.2009.09.003

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118. https://doi.org/10.1073/pnas.2024292118

Reimer, N. K., Kamble, S. V., Schmid, K., & Hewstone, M. (2022). Intergroup contact fosters more inclusive social identities. *Group Processes & Intergroup Relations*, *25*(1), 133–157. https://doi.org/10.1177/1368430220960795

Reimer, N. K., Schmid, K., Hewstone, M., & Al Ramiah, A. (2020). Self-categorization and social identification: Making sense of us and them. In D. Chadee (Ed.), *Theories in Social Psychology* (2nd ed.). Wiley-Blackwell.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, 64–99.

Richins, M. T., Barreto, M., Karl, A., & Lawrence, N. (2019). Empathic responses are reduced to competitive but not non-competitive outgroups. *Social Neuroscience*, *14*(3), 345-358. https://doi.org/10.1080/17470919.2018.1463927

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, *49*(23), 2856–2861. https://doi.org/10.1016/j.visres.2009.08.025

Riečanský, I., Paul, N., Kölble, S., Stieger, S., & Lamm, C. (2015). Beta oscillations reveal ethnicity ingroup bias in sensorimotor resonance to pain of others. *Social Cognitive and Affective Neuroscience, 10*(7), 893–901. https://doi.org/10.1093/scan/nsu139

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309–338. https://doi.org/10.1037/met0000184

Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, *27*(22), 1200–1206. https://doi.org/10.1016/j.cub.2017.09.060

Roberts, G., Raihani, N., Bshary, R., Manrique, H. M., Farina, A., Samu, F., & Barclay, P. (2021). The benefits of being seen to help others: Indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1838), 20200290. https://doi.org/10.1098/rstb.2020.0290

Romano, A., & Balliet, D. (2017). Reciprocity outperforms conformity to promote cooperation. *Psychological Science*, *28*(10), 1490–1502. https://doi.org/10.1177/0956797617714828

Rothmaler, K., Berger, P., & Wiesmann, C. G. (2022). Timing matters: Disentangling the neurocognitive sequence of mentalizing. *Trends in Cognitive Sciences*, *26*(11), 906–908. https://doi.org/10.1016/j.tics.2022.09.001

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., ... & Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour, 5*(10), 1369–1380. https://doi.org/10.1038/s41562-021-01092-x

Saguy, T., Szekeres, H., Nouri, R., Goldenberg, A., Doron, G., Dovidio, J. F., ... & Halperin, E. (2015). Awareness of intergroup help can rehumanize the out-group. *Social*

Psychological and Personality Science, *6*(5), 551–558. https://doi.org/10.1177/1948550615574748

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications*, 71–78. https://doi.org/10.1145/355017.355028

Santos, L. A., Voelkel, J. G., Willer, R., & Zaki, J. (2022). Belief in the Utility of Cross-Partisan Empathy Reduces Partisan Animosity and Facilitates Political Persuasion. *Psychological Science*, *33*(9), 1557–1573. https://doi.org/10.1177/09567976221098594

Scheffer, J. A., Cameron, C. D., & Inzlicht, M. (2022). Caring Is Costly: People Avoid the Cognitive Work of Compassion. *Journal of Experimental Psychology: General*, *151*(1), 172–196. https://doi.org/10.1037/xge0001073

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic Features of Emotional Expressions Are Processed Preferentially. *PLOS ONE*, *7*(7), e41792. https://doi.org/10.1371/journal.pone.0041792

Schiffmann, R., & Wicklund, R. A. (1992). The Minimal Group Paradigm and its Minimal Psychology: On Equating Social Identity with Arbitrary Group Membership. *Theory & Psychology*, *2*(1), 29–50. https://doi.org/10.1177/0959354392021002

Schnell, K., Bluschke, S., Konradt, B., & Walter, H. (2011). Functional relations of empathy and mentalizing: An fMRI study on the neural basis of cognitive empathy. *NeuroImage*, *54*(2), 1743–1754. https://doi.org/10.1016/j.neuroimage.2010.08.024

Schmid, P. C., Schmid Mast, M., Bombari, D., Mast, F. W., & Lobmaier, J. S. (2011). How Mood States Affect Information Processing During Facial Emotion Recognition: An

Eye Tracking Study. *Swiss Journal of Psychology*, *70*(4), 223–231. https://doi.org/10.1024/1421-0185/a000060

Schoofs, L., Claeys, A.-S., De Waele, A., & Cauberghe, V. (2019). The role of empathy in crisis communication: Providing a deeper understanding of how organizational crises and crisis communication affect reputation. *Public Relations Review*, *45*(5), 101851. https://doi.org/10.1016/j.pubrev.2019.101851

Schumann, A., Kietzer, S., Ebel, J., & Bär, K. J. (2020). Sympathetic and Parasympathetic Modulation of Pupillary Unrest. *Frontiers in Neuroscience*, *14*, 178–189. https://www.frontiersin.org/articles/10.3389/fnins.2020.00178

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology*, *107*(3), 475–493. https://doi.org/10.1037/a0036738

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision*, *14*(13), 14–25. https://doi.org/10.1167/14.13.14

Scott, W., & Cherrington, D. (1974). Effects of competitive, cooperative, and individualistic reinforcement contingencies. *Journal of Personality and Social Psychology*, *30*, 748–758. https://doi.org/10.1037/h0037534

Selbing, I., Lindström, B., & Olsson, A. (2014). Demonstrator skill modulates observational aversive learning. *Cognition*, *133*(1), 128–139. https://doi.org/10.1016/j.cognition.2014.06.010

Sen, M., & Wasow, O. (2016). Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics. *Annual Review of Political Science*, *19*(1), 499–522. https://doi.org/10.1146/annurev-polisci-032015-010015

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi.org/10.1093/brain/awn279

Sharot, T., & Garrett, N. (2016). Forming Beliefs: Why Valence Matters. *Trends in Cognitive Sciences*, *20*(1), 25–33. https://doi.org/10.1016/j.tics.2015.11.002

Sheng, F., & Han, S. (2012). Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *NeuroImage*, *61*(4), 786–797. https://doi.org/10.1016/j.neuroimage.2012.04.028

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R. & Sherif, C. W. (1961). *Intergroup cooperation and competition: The Robbers Cave experiment*. University Book Exchange.

Sherman, D. K., & Cohen, G. L. (2006). The Psychology of Self-defense: Self-Affirmation Theory. *Advances in Experimental Social Psychology*, *38*, 183–242. https://doi.org/10.1016/S0065-2601(06)38004-5

Shnabel, N., Halabi, S., & SimanTov-Nachlieli, I. (2015). Group apology under unstable status relations: Perceptions of insincerity hinder reconciliation and forgiveness. *Group Processes & Intergroup Relations*, *18*(5), 716–725. https://doi.org/10.1177/1368430214546069

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321. https://doi.org/10.3758/BF03210755

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*(18), 875–878. https://doi.org/10.1016/j.cub.2014.06.054

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for Pain Involves the Affective but not Sensory Components of Pain. *Science*, *303*(5661), 1157–1162. https://doi.org/10.1126/science.1093535

Singh, B., Mellinger, C., Earls, H. A., Tran, J., Bardsley, B., & Correll, J. (2022). Does Cross-Race Contact Improve Cross-Race Face Perception? A Meta-Analysis of the Cross-Race Deficit and Contact. *Personality and Social Psychology Bulletin*, *48*(6), 865–887. https://doi.org/10.1177/01461672211024463

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & Emotion*, *7*(3–4), 233–269. https://doi.org/10.1080/02699939308409189

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Smith, E. R., & Mackie, D. M. (2015). Dynamics of Group-Based Emotions: Insights From Intergroup Emotions Theory. *Emotion Review*, *7*(4), 349–354. https://doi.org/10.1177/1754073915590614

Smith, R. H., Powell, C. A. J., Combs, D. J. Y., & Schurtz, D. R. (2009). Exploring the When and Why of Schadenfreude. *Social and Personality Psychology Compass*, *3*(4), 530–546. https://doi.org/10.1111/j.1751-9004.2009.00181.x

Smith, R. H., & van Dijk, W. W. (2018). Schadenfreude and Gluckschmerz. *Emotion Review*, *10*(4), 293–304. https://doi.org/10.1177/1754073918765657

Soble, J., Spanierman, L., & Liao, H.-Y. (2011). Effects of a Brief Video Intervention on White University Students' Racial Attitudes. *Journal of Counseling Psychology*, *58*, 151–157. https://doi.org/10.1037/a0021158

Sparks, E., Schinkel, M. G., & Moore, C. (2017). Affiliation affects generosity in young children: The roles of minimal group membership and shared interests. *Journal of Experimental Child Psychology*, *159*, 242–262. https://doi.org/10.1016/j.jecp.2017.02.007

Statistical Office of the Republic of Serbia. (2018). *Demographic Yearbook*. Retrieved March 22, 2022, from https://www.stat.gov.rs/en-us/publikacije/publication/?p=12244

statmerkur (https://stats.stackexchange.com/users/136579/statmerkur, 2019, December 8). *Compute partial eta squared for all fixed effects ANOVAs from a lme4 model*. https://stats.stackexchange.com/q/439842

Staub, E., Pearlman, L. A., Gubin, A., & Hagengimana, A. (2005). Healing, Reconciliation, Forgiving and the Prevention of Violence after Genocide or Mass Killing: An Intervention and Its Experimental Evaluation in Rwanda. *Journal of Social and Clinical Psychology*, *24*(3), 297–334. https://doi.org/10.1521/jscp.24.3.297.65617

Staunton, M., Louis, W. R., Smith, J. R., Terry, D. J., & McDonald, R. I. (2014). How negative descriptive norms for healthy eating undermine the effects of positive

injunctive norms. *Journal of Applied Social Psychology*, *44*(4), 319–330. https://doi.org/10.1111/jasp.12223

Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, *115*(49), 12435–12440. https://doi.org/10.1073/pnas.1803470115

Stemerding, L. E., van Ast, V. A., Gerlicher, A. M. V., & Kindt, M. (2022). Pupil dilation and skin conductance as measures of prediction error in aversive learning. *Behaviour Research and Therapy*, *157*, 104164. https://doi.org/10.1016/j.brat.2022.104164

Stets, J. E., & Burke, P. J. (2000). Identity Theory and Social Identity Theory. *Social Psychology Quarterly*, *63*(3), 224–237. https://doi.org/10.2307/2695870

Stevens, F., & Taber, K. (2021). The neuroscience of empathy and compassion in pro-social behavior. *Neuropsychologia*, *159*, 107925. https://doi.org/10.1016/j.neuropsychologia.2021.107925

Stillerman, B., Lindström, B., Schultner, D., Hackel, L. M., Hagen, D., Jostmann, N., & Amodio, D. (2020). *Societal stereotypes shape learning to produce group-based preferences*. PsyArXiv. https://doi.org/10.31234/osf.io/mwztc

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*(6), 795–797. https://doi.org/10.1038/nn.4296

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial Image Manipulation: A Tool for Investigating Social Perception. *Social Psychological and Personality Science*, *8*(5), 538–551. https://doi.org/10.1177/1948550617697176

Tajfel, H. (1972a). Social categorization. In S. Moscovici (Ed.), *Introduction a la psychologie sociale* (Vol.1). Larouse.

Tajfel, H. (1972b). Experiments in a vacuum. In J. Israel & H. Tajfel (Eds.), *The context of social psychology* (pp. 69–119). Academic Press.

Tajfel, H. (Ed.) (1978). *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. Academic Press.

Tajfel, H. (1981). *Human groups and social categories*. Cambridge University Press.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33 – 47). Brooks/Cole.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude. *Science*, *323*(5916), 937–939. https://doi.org/10.1126/science.1165604

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, *10*(1), 181–211. https://doi.org/10.1111/sipr.12022

Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology*, *48*(3), 427–446. https://doi.org/10.1348/014466608X373589

Tarrant, M., & Hadert, A. (2010). Empathic Experience and Attitudes Toward Stigmatized Groups: Evidence for Attitude Generalization. *Journal of Applied Social Psychology*, *40*(7), 1635–1656. https://doi.org/10.1111/j.1559-1816.2010.00633.x

Teding van Berkhout, E., & Malouff, J. M. (2016). The efficacy of empathy training: A meta-analysis of randomized controlled trials. *Journal of Counseling Psychology*, *63*(1), 32–41. https://doi.org/10.1037/cou0000093

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic Games: An Introduction and Guide for Research. *Collabra: Psychology*, *7*(1), 19004. https://doi.org/10.1525/collabra.19004

Thomas, A. J., Woo, B., Nettle, D., Spelke, E., & Saxe, R. (2022). Early concepts of intimacy: Young humans use saliva sharing to infer close relationships. *Science*, *375*(6578), 311–315. 10.1126/science.abh1054

Thornton, M. A., & Tamir, D. I. (2022). Six dimensions describe action understanding: The ACT-FASTaxonomy. *Journal of Personality and Social Psychology, 122*, 577–605. https://doi.org/10.1037/pspa0000286

Thornton, M. A., Weaverdyck, M. E., Mildner, J. N., & Tamir, D. I. (2019). People represent their own mental states more distinctly than those of others. *Nature Communications, 10*(1), 1–9. https://doi.org/10.1038/s41467-019-10083-6

Thornton, M. A., Wolf, S., Reilly, B. J., Slingerland, E. G., & Tamir, D. I. (2022). The 3d Mind Model Characterizes How People Understand Mental States Across Modern and Historical Cultures. *Affective Science, 3*(1), 93–104. https://doi.org/10.1007/s42761-021-00089-z

To, C., Kilduff, G. J., Ordoñez, L., & Schweitzer, M. E. (2018). Going for it on Fourth Down: Rivalry Increases Risk Taking, Physiological Arousal, and Promotion Focus. *Academy of Management Journal, 61*(4), 1281–1306. https://doi.org/10.5465/amj.2016.0850

To, C., Kilduff, G., & Rosikiewicz, B. (2020). When Interpersonal Competition Helps and When It Harms: An Integration via Challenge and Threat. *Academy of Management Annals, 14*, 908–934. https://doi.org/10.5465/annals.2016.0145

Toma, C., Corneille, O., & Yzerbyt, V. (2012). Holding a Mirror Up to the Self: Egocentric Similarity Beliefs Underlie Social Projection in Cooperation. *Personality and Social Psychology Bulletin*, *38*(10), 1259–1271. https://doi.org/10.1177/0146167212449022

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology, 94*(3), 516–530. https://doi.org/10.1037/0022-3514.94.3.516

Tsikandilakis, M., Bali, P., Derrfuss, J., & Chapman, P. (2020). Anger and hostility: Are they different? An analytical exploration of facial-expressive differences, and physiological and facial-emotional responses. *Cognition and Emotion*, *34*(3), 581–595. https://doi.org/10.1080/02699931.2019.1664415

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the Social Group: A self-categorization Theory*. Blackwell.

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237–252. https://doi.org/10.1111/j.2044-8309.1986.tb00732.x

Turner, J. C., & Reynolds, K. J. (2010). The story of social identity. In *Rediscovering social identity: Key readings*. Psychology Press, Taylor & Francis.

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science*, *27*(8), 1109–1122. 10.1177/0956797616650875

United Nations International Criminal Tribunal for the Former Yugoslavia (UN-ICTY). *Croatia - 1991-1995*. Retrieved March 29, 2022, from https://www.icty.org/en/about/what-former-yugoslavia/conflicts

United Nations High Commissioner for Refugees. (2022). *Ukraine situation Flash Update #30*. https://data.unhcr.org/en/documents/details/95676

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

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion, 11*(2), 262–266. https://doi.org/10.1037/a0021302

Vallentin-Holbech, L., Rasmussen, B. M., & Stock, C. (2017). Are perceptions of social norms regarding peer alcohol and other drug use associated with personal use in Danish adolescents? *Scandinavian Journal of Public Health*, *45*(8), 757–764. https://doi.org/10.1177/1403494817724313

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, *25*(11), 913–916. https://doi.org/10.1016/j.tics.2021.07.013

Van Bavel, J. J., Swencionis, J. K., O'Connor, R. C., & Cunningham, W. A. (2012). Motivated social memory: Belonging needs moderate the own-group bias in face recognition. *Journal of Experimental Social Psychology*, *48*(3), 707–713. https://doi.org/10.1016/j.jesp.2012.01.006

Van Der Linden, S. (2015). Exploring beliefs about bottled water and intentions to reduce consumption: The dual-effect of social norm activation and persuasive information. *Environment and Behavior*, *47*(5), 526–550. https://doi.org/10.1177/0013916513515239

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. https://doi.org/10.1037/a0023853

van Diepen, P., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, *10*(1), 79–100. https://doi.org/10.1080/713756668

van Dijk, E., Molenmaker, W. E., & de Kwaadsteniet, E. W. (2015). Promoting cooperation in social dilemmas: The use of sanctions. *Current Opinion in Psychology*, *6*, 118–122. https://doi.org/10.1016/j.copsyc.2015.07.006

van Ham, M., & Tammaru, T. (2016). New perspectives on ethnic segregation over time and space. A domains approach. *Urban Geography*, *37*(7), 953–962. https://doi.org/10.1080/02723638.2016.114215

van Veelen, R., Otten, S., & Hansen, N. (2013). Social identification when an in-group identity is unclear: The role of self-anchoring and self-stereotyping. *British Journal of Social Psychology*, *52*(3), 543–562. https://doi.org/10.1111/j.2044-8309.2012.02110.x

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*(4), 455–472. https://doi.org/10.1002/ejsp.163

Vollberg, M. C., & Cikara, M. (2018). The neuroscience of intergroup emotion. *Current Opinion in Psychology*, *24*, 48–52. https://doi.org/10.1016/j.copsyc.2018.05.003

Vollberg, M. C., & Cikara, M. (2022). *Affective Prediction Errors in Persistence and Escalation of Aggression*. PsyArXiv. https://doi.org/10.31234/osf.io/59k3c

Vollberg, M. C., Gaesser, B., & Cikara, M. (2021). Activating episodic simulation increases affective empathy. *Cognition*, *209*, 104558. https://doi.org/10.1016/j.cognition.2020.104558

Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, *20*(4), 461–475. https://doi.org/10.1002/acp.1191

Wang, C.-A., Baird, T., Huang, J., Coutinho, J. D., Brien, D. C., & Munoz, D. P. (2018). Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Frontiers in Neurology*, *9*, 1029–1041. https://www.frontiersin.org/articles/10.3389/fneur.2018.01029

Wang, Y., Harris, P. L., Pei, M., & Su, Y. (2022). Do Bad People Deserve Empathy? Selective Empathy Based on Targets' Moral Characteristics. *Affective Science*, 1–16. https://doi.org/10.1007/s42761-022-00165-y

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

Ward, J., Schaal, M., Sullivan, J., Bowen, M., Erdmann, J., & Hojat, M. (2009). Reliability and Validity of the Jefferson Scale of Empathy in Undergraduate Nursing Students. *Journal of Nursing Measurement*, *17*, 73–88. https://doi.org/10.1891/1061-3749.17.1.73

Watanabe, S., Weiner, D. S., & Laurent, S. M. (2022). Schadenfreude for undeserved misfortunes: The unexpected consequences of endorsing a strong belief in a just world. *Journal of Experimental Social Psychology*, *101*, 104336. https://doi.org/10.1016/j.jesp.2022.104336

Waytz, A. (2016). The limits of empathy. *Harvard Business Review*, *94*(1), 68–73.

Waytz, A., Young, L. L., & Ginges, J. (2014). Motive attribution asymmetry for love vs. hate drives intractable conflict. *Proceedings of the National Academy of Sciences*, *111*(44), 15687–15692. https://doi.org/10.1073/pnas.1414146111

Weisel, O., & Böhm, R. (2015). “Ingroup love” and “outgroup hate” in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, *60*, 110–120. https://doi.org/10.1016/j.jesp.2015.04.008

Weisz, E., Chen, P., Ong, D. C., Carlson, R. W., Clark, M. D., & Zaki, J. (2022). A brief intervention to motivate empathy among middle school students. *Journal of Experimental Psychology: General*, *151*(12), 3144–3153. https://doi.org/10.1037/xge0001249

Weisz, E., & Cikara, M. (2021). Strategic Regulation of Empathy. *Trends in Cognitive Sciences*, *25*(3), 213–227. https://doi.org/10.1016/j.tics.2020.12.002

Weisz, E., Ong, D. C., Carlson, R. W., & Zaki, J. (2021). Building empathy through motivation-based interventions. *Emotion*, *21*(5), 990–999. https://doi.org/10.1037/emo0000929

Weisz, E., & Zaki, J. (2017). Empathy-building interventions: A review of existing work and suggestions for future directions. In M. J. Doty, E. Seppala, E. Simon-Thomas, D. Cameron, & S. Brown (Eds.), *The Oxford handbook of compassion science* (pp. 205–217). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190464684.001.0001

Weisz, E., & Zaki, J. (2018). Motivated empathy: A social neuroscience perspective. *Current Opinion in Psychology*, *24*, 67–71. https://doi.org/10.1016/j.copsyc.2018.05.005

Wickline, V. B., Bailey, W., & Nowicki, S. (2009). Cultural In-Group Advantage: Emotion Recognition in African American and European American Faces and Voices. *The Journal of Genetic Psychology*, *170*(1), 5–30. https://doi.org/10.3200/GNTP.170.1.5-30

Wilder, D. A. (1984). Empirical contributions: Predictions of belief homogeneity and similarity following social categorization. *British Journal of Social Psychology*, *23*(4), 323–333. https://doi.org/10.1111/j.2044-8309.1984.tb00648.x

Willems, A. L., & Vervliet, B. (2021). When nothing matters: Assessing markers of expectancy violation during omissions of threat. *Behaviour Research and Therapy*, *136*, 103764. https://doi.org/10.1016/j.brat.2020.103764

Willenbockel, V., Sadr, J., Fiset, D., Horne, G., Gosselin, F., & Tanaka, J. (2010). The SHINE toolbox for controlling low-level image properties. *Journal of Vision*, *10*(7), 653–653. https://doi.org/10.1167/10.7.653

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, *50*, 314–321. https://doi.org/10.1037/0022-3514.50.2.314

Wood, W., & Neal, D. (2007). A New Look at Habits and the Habit-Goal Interface. *Psychological Review*, *114*, 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*, 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Reputation, Gossip, and Human Cooperation. *Social and Personality Psychology Compass*, *10*(6), 350–364. https://doi.org/10.1111/spc3.12255

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do You Feel My Pain? Racial Group Membership Modulates Empathic Neural Responses. *Journal of Neuroscience*, *29*(26), 8525–8529. https://doi.org/10.1523/JNEUROSCI.2418-09.2009

Yamagishi, T., & Kiyonari, T. (2000). The Group as the Container of Generalized Reciprocity. *Social Psychology Quarterly*, *63*(2), 116–132. https://doi.org/10.2307/2695887

Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: In-group love or out-group hate? *Evolution and Human Behavior*, *30*(4), 229–237. https://doi.org/10.1016/j.evolhumbehav.2009.02.004

Yamagishi, T., Mifune, N., Liu, J. H., & Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner's dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology*, *11*(3), 196–207. https://doi.org/10.1111/j.1467-839X.2008.00258.x

Yang, X., & Dunham, Y. (2019). Minimal but meaningful: Probing the limits of randomly assigned social identities. *Journal of Experimental Child Psychology*, *185*, 19–34. https://doi.org/10.1016/j.jecp.2019.04.013

Yeager, D. S., & Dweck, C. S. (2020). What Can Be Learned from Growth Mindset Controversies? *The American Psychologist*, *75*(9), 1269–1284. https://doi.org/10.1037/amp0000794

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology*, *99*(6), 964–977. https://doi.org/10.1037/a0020400

Young, S. G., & Wilson, J. P. (2018). A minimal ingroup advantage in emotion identification confidence. *Cognition and Emotion*, *32*(1), 192–199. https://doi.org/10.1080/02699931.2016.1273199

Yrttiaho, S., Niehaus, D., Thomas, E., & Leppänen, J. M. (2017). Mothers' pupillary responses to infant facial expressions. *Behavioral and Brain Functions*, *13*(1), 1–12. https://doi.org/10.1186/s12993-017-0120-9

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*(6), 1608–1647. https://doi.org/10.1037/a0037679

Zaki, J. (2020). Catastrophe Compassion: Understanding and Extending Prosociality Under Crisis. *Trends in Cognitive Sciences*, *24*(8), 587–589. https://doi.org/10.1016/j.tics.2020.05.006

Zaki, J., & Cikara, M. (2015). Addressing Empathic Failures. *Current Directions in Psychological Science*, *24*(6), 471–476. https://doi.org/10.1177/0963721415599978

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, *15*(5), 675–680. https://doi.org/10.1038/nn.3085

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social Influence Modulates the Neural Computation of Value. *Psychological Science*, *22*(7), 894–900. https://doi.org/10.1177/0956797611411057

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*(2), 175–189. https://doi.org/10.1037/a0017990

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences*, *115*(19), 4532–4540. https://doi.org/10.1073/pnas.1708960115

APPENDICES

CHAPTER 2: SUPPLEMENTARY MATERIALS

APPENDIX 2.1: Relationship between SCR metrics and valence judgments of facial expressions (Experiment 3)

Before running any exploratory analyses to examine the correlations between SCR metrics and judgments of facial expressions, I checked whether the results from Experiment 3's reduced sample reflected the whole-sample results reported in Experiment 1. Specifically, I checked whether group category membership of the expresser and coalitional relations with the expresser influenced lab-based participants' valence, but not arousal and genuineness judgments of facial expressions. In the domain of valence judgments, I observed a significant main effect of coalitional relations with the expresser ($F(1, 3475.8) = 6.72$, $p = .009$) and a significant main effect of expressers' group category membership ($F(1, 3474.9) = 6.16$, $p = .013$) alongside a significant main effect of emotion category ($F(2, 34.8) = 137.69$, $p < .001$). This pattern aligns with the whole-sample results from Experiment 1. Furthermore, in line with the full results of Experiment 1, no effects other than the main effect of emotion category (arousal: $F(2, 34.4) = 42.12$, $p < .001$; genuineness: $F(2, 25.52) = 50.69$, $p < .001$) were found in the domain of arousal (all $ps > .155$) and genuineness judgments (all $ps > .468$)[82].

I proceeded to examine the following: (a) whether participants' SCR during the presentation of the textual cue correlated with their valence judgment[83] on the same trial, (b)

[82] These analyses were run only to check whether lab-based participants' judgments broadly track whole-sample patterns reported in Experiment 1. Therefore, any resulting estimates should not be independently interpreted.

[83] Valence judgments were standardised within each participant and emotion category given that the SCR observed during the textual cue could only index participants' responsivity to group category membership and coalitional relations.

whether participants' SCR during the presentation of the facial expression correlated with their valence judgment on the same trial, and finally, (c) whether the extent to which participants' SCR discriminated between minimal ingroup and outgroup members (i.e., $\Delta SCR_{GROUP} = SCR_{INGROUP} - SCR_{OUTGROUP}$) and between those with whom one was to compete versus cooperate (i.e., $\Delta SCR_{COALITION} = SCR_{COMPETITION} - SCR_{COOPERATION}$) predicted the extent to which these factors influenced participants' valence judgments of facial expressions (i.e., $\Delta Valence_{GROUP} = Valence_{INGROUP} - Valence_{OUTGROUP}$; $\Delta Valence_{COALITION} = Valence_{COMPETITION} - Valence_{COOPERATION}$).

Under (a) and (b), I observed no significant repeated-measures correlations between participants' trial-by-trial SCR metrics, derived either during the presentation of the textual cue or the presentation of the facial expression, and their subsequent valence judgments (textual cues: all $|rs| < .01$, all $ps > .367$; facial expressions: all $|rs| < .02$, all $ps > .194$). Furthermore, no significant repeated-measures correlations were observed even when I focused on the first session of the emotion task (i.e., the first 32 trials) to account for the fact that SCR habituates over time (textual cues: all $|rs| < .05$, all $ps > .090$; facial expressions: all $|rs| < .04$, all $ps > .130$). Under (c), I only observed a significant positive correlation between how much participants' TTP-identified SCR amplitudes discriminated between "competitive" and "cooperative" textual cues and how much participants discriminated between those with whom they were to compete versus cooperate in their valence judgments of facial expressions ($r = .36$, $p = .036$). Specifically, participants who displayed stronger SCR to anticipated competition than cooperation mostly did not discriminate on the basis of coalitional relations in their valence judgments, while participants who displayed stronger SCR to anticipated cooperation than competition mostly showed a cooperation positivity bias. However, this correlation became insignificant when a single flagged outlier participant was excluded ($r = .30$, $p = .085$). Furthermore, this correlation was specific to TTP-identified

SCR amplitudes and was not observed using other SCR metrics (CDA-identified SCR amplitudes: $r = .20$, $p = .258$; CDA-identified phasic driver: $r = .22$, $p = .204$). Finally, no significant correlations were observed when SCR metrics during the presentation of the facial expression were used (all $rs < .32$, all $ps > .068$).

CHAPTER 3: SUPPLEMENTARY MATERIALS

APPENDIX 3.1: Male first names (Experiment 4)

Table A3.1.1

List of Croatian names used in Experiment 4

	Name categorisation						
	Bosniak	Croat	Macedonian	Serb	Slovene	Multiple groups	None of the groups
Antun	1	42	0	0	0	3	0
Domagoj	0	34	0	0	3	6	3
Franjo	1	41	0	0	1	4	0
Hrvoje	0	43	0	1	0	2	0
Juraj	1	42	0	0	0	4	0
Mate	1	43	0	0	1	2	0
Stipe	0	43	1	1	0	2	0
Šime	1	42	1	0	1	2	0

Notes. This table contains results from a pilot study (N = 47, 22 Croatian, 25 Serbian, 11 men, 36 women, M_{AGE} = 26.30, SD_{AGE} = 6.07) where participants were asked to guess the ethnicity of the target based only on their first name. Numbers indicate how many participants said the target was of a particular ethnicity.

Table A3.1.2

List of Serbian names used in Experiment 4

	Name categorisation						
	Bosniak	Croat	Macedonian	Serb	Slovene	Multiple groups	None of the groups
Đorđe	2	1	1	32	0	10	0
Jovan	0	0	0	37	1	9	0
Mihajlo	1	0	1	39	2	4	0
Miloš	0	0	1	40	1	5	0
Nebojša	2	0	0	41	0	4	0
Nemanja	1	0	0	43	0	3	0
Sava	1	0	0	36	1	9	0
Vukašin	0	0	3	39	2	3	0

Notes. This table contains results from a pilot study (N = 47, 22 Croatian, 25 Serbian, 11 men, 36 women, M_{AGE} = 26.30, SD_{AGE} = 6.07) where participants were asked to guess the ethnicity of the target based only on their first name. Numbers indicate how many participants said the target was of a particular ethnicity.

APPENDIX 3.2: Pilot studies for stimuli selection (Experiments 4, 5 & 6)

Experiment 4

Final 16 stimuli included in Experiment 4 were chosen based on a pilot study (N = 62; 32 Croatian, 30 Serbian; 12 men, 50 women; M_{AGE} = 27.55, SD_{AGE} = 9.31). For the pilot study, 43 stories with no group labels were created out of which I selected eight pairs of stories that satisfied the following criteria: (a) stories in the same pair had the same dominant emotion, (b) stories in the same pair were as equal as possible in perceived deservingness of the experience, and (c) stories in the same pair were as equal as possible in perceived valence of the experience. Stories from each matched pair were allocated to two different sets, with stories from each set labelled as ingroup and outgroup members interchangeably in Experiment 4. In Experiment 4, four dominant emotions were represented across stimuli: anger/annoyance, fear, disappointment/sadness, guilt/shame. I did not have any hypotheses related to the four dominant emotions; they were just included to represent a wide variety of real-life emotional experiences.

The selection of stories was done separately in the Croatian and Serbian sample until a combination of stories that satisfied all three criteria in both samples was found. As a final check, a whole sample analysis was run to see if there were any differences between the two samples as well as between the two sets of stimuli. Two-way ANOVA with sample and stimuli set as variables of interest revealed no differences between the two sets in valence (M_{SET1} = 84.65; M_{SET2} = 85.55; p = .227) and deservingness (M_{SET1} = 26.96; M_{SET2} = 27.23; p = .883). No additional differences between the two samples (all ps > .248) and no interaction effects (sample x set) were found (all ps > .316).

Experiment 5

For Experiment 5, final 16 stimuli were also chosen based on a pilot study (N = 29; 12 Leave, 17 Remain; 11 men, 18 women; M_{AGE} = 33.90, SD_{AGE} = 9.33). This pilot study included 42 stories with no group labels and followed the same structure and criteria as the pilot study for Experiment 4. After selecting four sets of stimuli (two sets of four group-irrelevant stimuli, two sets of four group-relevant stimuli), I ran a two-way ANOVA with type of the emotional experience and stimuli set as variables of interest to examine whether there are any significant differences in perceived deservingness and valence. No significant differences were found (all *ps* > .108; deservingness of group-irrelevant experiences: M_{SET1} = 20.55; M_{SET2} = 21.34; deservingness of group-relevant experiences: M_{SET1} = 22.33; M_{SET2} = 18.31; valence of group-irrelevant experiences: M_{SET1} = 81.99; M_{SET2} = 82.14; valence of group-relevant experiences: M_{SET1} = 82.17; M_{SET2} = 81.60). In Experiment 5, two dominant emotions were represented across stimuli: anger/annoyance and disappointment/sadness.

Experiment 6

To select the stimuli for Experiment 6, a third pilot study (N = 20; 6 men, 14 women; M_{AGE} = 24.15, SD_{AGE} = 12.24), following the same structure and criteria as the pilot studies in previous experiments, was conducted. From the original pool of 46 stimuli, I selected two sets of eight stimuli which were labelled as belonging to minimal ingroup and outgroup members interchangeably in Experiment 6. At the end, I ran a whole-sample analysis to test whether there are any differences in perceived deservingness and valence, but no such differences were observed (all *ps* > .789; deservingness: M_{SET1} = 32.11; M_{SET2} = 32.70; valence: M_{SET1} = 79.49; M_{SET2} = 79.76). In Experiment 6, three dominant emotions were represented across stimuli: anger/annoyance, disappointment/sadness, and guilt.

APPENDIX 3.3: Model building and selection – Prejudice perceptions (Experiments 4 & 5)

Experiment 4

Table A3.3.1

Model building and comparison for the average ingroup empathiser condition (Experiment 4)

Sampling units		N observations = 11008 N participants = 344; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x NT	Baseline	-	-	Intercepts, group slope	Intercepts	96144	96210	-48063	9		
Group x NT x InIn	InIn	Baseline	InIn (three-way interaction)	Same	Same	96121	96215	-48047	13	4	31.60

Group x NT x InOut	InOut	Baseline	InOut (three-way interaction)	Same	Same	96042	96137	-48008	13	4	109.77
Group x NT x InIn + Group x NT x InOut	InIn + InOut	InOut	InIn, InOut (two three-way interactions)	Same	Same	96017	96141	-47992	17	4	33.11
Group x NT x InIn x InOut	InIn x InOut	InIn + InOut	InIn x InOut (four-way interaction)	Same	Same	96018	96172	-47998	21	4	6.76

Notes. Group = Group category membership of the target (own-group, other-group). NT = Norm type (emotion perception, empathy). InIn = Expected prejudice of ingroup members towards ingroup targets. InOut = Expected prejudice of ingroup members towards outgroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the selected model whose output, after controlling for correlated prejudice perceptions (OutOut, OutIn), is reported in the main text.

Table A3.3.2

Model building and comparison for the average outgroup empathiser condition (Experiment 4)

Sampling units		N observations = 11008 N participants = 344; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x NT	Baseline	-	-	Intercepts, group slope	Intercepts	95565	95631	-47774	9		
Group x NT x OutOut	OutOut	Baseline	OutOut (three-way interaction)	Same	Same	95550	95645	-47762	13	4	23.27

Group x NT x OutIn	OutIn	Baseline	OutIn (three-way interaction)	Same	Same	95458	95553	-47716	13	4	115.01
Group x NT x OutOut + Group x NT x OutIn	OutOut + OutIn	OutIn	OutOut, OutIn (two three-way interactions)	Same	Same	95449	95573	-47707	17	4	17.5
Group x NT x OutOut x OutIn	OutOut x OutIn	OutIn	OutOut x OutIn (four-way interaction)	Same	Same	95455	95608	-47707	21	8	19.29

Notes. Group = Group category membership of the target (ingroup, outgroup). NT = Norm type (emotion perception, empathy). OutOut = Expected prejudice of outgroup members towards outgroup targets. OutIn = Expected prejudice of outgroup members towards ingroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the selected model whose output, after controlling for correlated prejudice perceptions (InOut), is reported in the main text.

Experiment 5

Table A3.3.3

Model building and comparison for the average ingroup empathiser condition – Empathy responses (Experiment 5)

Sampling units		N observations = 3103 N participants = 195; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x ET	Baseline	-	-	Intercepts, group and ET slopes	Intercepts	26838	26911	-13407	12		
Group x ET x InIn	InIn	Baseline	InIn (three-way interaction)	Same	Same	26839	26936	-13404	16	4	7.26

Group x ET x InOut	InOut	Baseline	InOut (three-way interaction)	Same	Same	26830	26927	-13339	16	4	16.57
Group x ET x InIn + Group x ET x InOut	InIn + InOut	InOut	InIn, InOut (two three-way interactions)	Same	Same	26831	26952	-13395	20	4	7.05
Group x ET x InIn x InOut	InIn x InOut	InOut	InIn x InOut (four-way interaction)	Same	Same	26835	26980	-13393	24	8	11.23

Notes. Group = Group category membership of the target (ingroup, outgroup). ET = Experience type (group-irrelevant, group-relevant). InIn = Expected prejudice of ingroup members towards ingroup targets. InOut = Expected prejudice of ingroup members towards outgroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the selected model whose output is reported in the main text.

Table A3.3.4

Model building and comparison for the average outgroup empathiser condition – Empathy responses (Experiment 5)

Sampling units		N observations = 3103 N participants = 195; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x ET	Baseline	-	-	Intercepts, group and ET slopes	Intercepts	27303	27375	-13369	12		
Group x ET x OutOut	OutOut	Baseline	OutOut (three-way interaction)	Same	Same	27297	27394	-13633	16	4	13.44

Group x ET x OutIn	OutIn	Baseline	OutIn (three-way interaction)	Same	Same	27279	27376	-13624	16	4	31.45
Group x ET x OutOut + Group x ET x OutIn	OutOut + OutIn	OutIn	OutOut, OutIn (two three-way interactions)	Same	Same	27281	27402	-13621	20	4	6.10
Group x ET x OutOut x OutIn	OutOut x OutIn	OutIn	OutOut x OutIn (four-way interaction)	Same	Same	27286	27431	-13619	24	8	9.29

Notes. Group = Group category membership of the target (ingroup, outgroup). ET = Experience type (group-irrelevant, group-relevant). OutOut = Expected prejudice of outgroup members towards outgroup targets. OutIn = Expected prejudice of outgroup members towards ingroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the selected model whose output is reported in the main text.

Table A3.3.5

Model building and comparison for the average ingroup empathiser condition – Emotion perception responses (Experiment 5)

Sampling units		N observations = 3103 N participants = 195; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x ET	Baseline	-	-	Intercepts, group and ET slopes	Intercepts	26344	26417	-13160	12		
Group x ET x InIn	InIn	Baseline	InIn (three-way interaction)	Same	Same	26339	26436	-13153	16	4	13.61

Group x ET x InOut	InOut	Baseline	InOut (three-way interaction)	Same	Same	26350	26447	-13159	16	4	2.16
Group x ET x InIn + Group x ET x InOut	InIn + InOut	InIn	InIn, InOut (two three-way interactions)	Same	Same	26339	26460	-13150	20	4	7.72
Group x ET x InIn x InOut	InIn x InOut	InIn	InIn x InOut (four-way interaction)	Same	Same	26343	26488	-13147	24	8	12.03

Notes. Group = Group category membership of the target (ingroup, outgroup). ET = Experience type (group-irrelevant, group-relevant). InIn = Expected prejudice of ingroup members towards ingroup targets. InOut = Expected prejudice of ingroup members towards outgroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the best performing model, but it did not include any significant interactions between prejudice perceptions and other factors of interest (group, experience type).

Table A3.3.6

Model building and comparison for the average outgroup empathiser condition – Emotion perception responses (Experiment 5)

Sampling units		N observations = 3103 N participants = 195; N stimuli = 16									
Model specification	**Model name**	**Best-performing nested alternative**	**Fixed effects added**	**Random effects**		**Model fit**				**LRT against nested**	
				Subjects	**Stimuli**	**AIC**	**BIC**	**LL**	**df**	**Δdf**	X^2
Group x ET	Baseline	-	-	Intercepts, group and ET slopes	Intercepts	26891	26963	-13433	12		
Group x ET x OutOut	OutOut	Baseline	OutOut (three-way interaction)	Same	Same	26891	26987	-13429	16	4	8.14

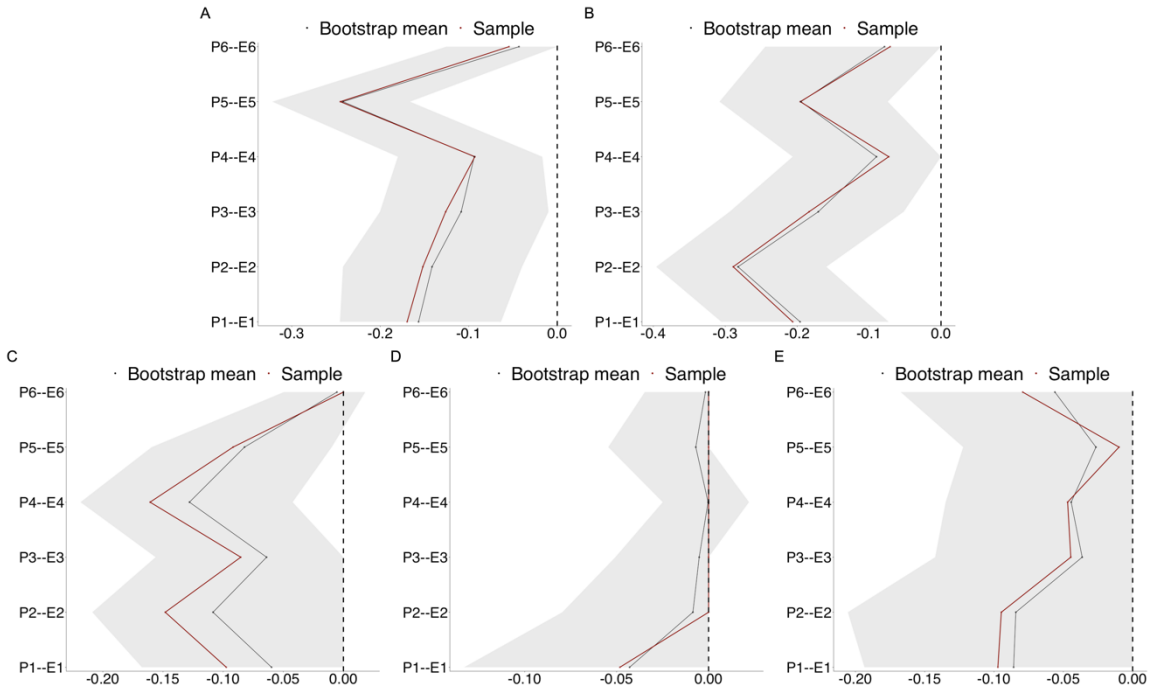
Group x ET x OutIn	OutIn	Baseline	OutIn (three-way interaction)	Same	Same	26889	26986	-13429	16	4	9.36
Group x ET x OutOut + Group x ET x OutIn	OutOut + OutIn	Baseline	OutOut, OutIn (two three-way interactions)	Same	Same	26888	27009	-13424	20	8	18.47
Group x ET x OutOut x OutIn	OutOut x OutIn	Baseline	OutOut x OutIn (four-way interaction)	Same	Same	26889	27034	-13420	24	12	26.01

Notes. Group = Group category membership of the target (ingroup, outgroup). ET = Experience type (group-irrelevant, group-relevant). OutOut = Expected prejudice of outgroup members towards outgroup targets. OutIn = Expected prejudice of outgroup members towards ingroup targets. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LL = Log-likelihood. df = Degrees of freedom. X^2 = Chi-square. LRT = Likelihood-ratio test. Model marked in bold is the best performing model.

APPENDIX 3.4: Accuracy of estimated network connections (Experiments 4 & 5)

Figure A3.4.1

Bootstrapped confidence intervals of relevant edge-weights for the estimated networks of prejudice and empathic norm perceptions



Notes. Sub-figures A, B, C, D and E correspond to networks 3.3A, 3.6A (acontextualised perceptions of empathic norms), 3.3B, 3.6B, and 3.6C (contextualised perceptions of empathic norms) in the main text of Chapter 3. E1 = Perceptions of ingroup members as empathic (A & B) or one's own empathy for ingroup targets (C-E). E2 = Perceptions of outgroup members as empathic (A & B) or one's own empathy for outgroup targets (C-E). E3 = Expected empathy of ingroup members for ingroup targets. E4 = Expected empathy of ingroup members for outgroup targets. E5 = Expected empathy of outgroup members for ingroup targets. E6 = Expected empathy of outgroup members for outgroup targets. P1 = One's own prejudice towards ingroup targets. P2 = One's own prejudice towards outgroup targets. P3 = Expected prejudice of ingroup members towards ingroup targets. P4 = Expected prejudice of ingroup members towards outgroup targets. P5 = Expected prejudice of outgroup members towards ingroup targets. P6 = Expected prejudice of outgroup members towards outgroup targets. Please note that the displayed confidence intervals should not be treated as a separate significance test. As LASSO regularisation was used for edge selection and estimation, all edges shown in Figures 3.3 and 3.6 in the main text of Chapter 3 were already deemed strong enough to be included in the model since LASSO shrinks all unnecessary edges to zero. Instead of using them as a significance test, confidence intervals should be used to (a) get a sense of edge weight precision and (b) compare edges to one another (Epskamp et al., 2018).

APPENDIX 3.5: Comparison of expected intergroup empathy bias (Experiments 4, 5 & 6)

Table A3.5.1

Comparison of acontextualised intergroup empathy bias participants expected from ingroup and outgroup members across experiments

Contrast	Empathiser	ΔIEB	95% confidence interval	t	p-value
Experiment 4 – Experiment 5	Average ingroup member	1.19	[-3.95, 6.33]	0.56	0.579
Experiment 4 – Experiment 6a	Average ingroup member	14.90	[8.78, 21.02]	5.84	< .001
Experiment 5 – Experiment 6a	Average ingroup member	13.71	[7.02, 20.4]	4.92	< .001
Experiment 4 – Experiment 5	Average outgroup member	-3.14	[-8.27, 2.00]	-1.46	.144

Experiment 4 – Experiment 6a	Average outgroup member	16.20	[10.08, 22.31]	6.35	< .001
Experiment 5 – Experiment 6a	Average outgroup member	19.33	[12.65, 26.02]	6.94	< .001

Notes. ΔIEB = Difference between the intergroup empathy bias participants expected from a particular empathiser (average ingroup member, average outgroup member) in the two compared experiments.

CHAPTER 4: SUPPLEMENTARY MATERIALS

APPENDIX 4.1: Fitted reinforcement learning (RL) models

In the first step, all RL models built on the Rescorla-Wagner learning rule:

$$Q_i{}^t = Q_i{}^{t-n} + \alpha * (EF_i{}^{t-n} - Q_i{}^{t-n}) \quad [1]$$

In equation 1, $Q_i{}^t$ is the perception of the intergroup empathic norm *i* (i.e., ingroup or outgroup intergroup empathic norm) in trial *t*, $Q_i{}^{t-n}$ is the preceding perception of the same intergroup empathic norm *i* in trial $t - n$, $EF_i{}^{t-n}$ is the empathic feedback related to intergroup empathic norm *i* received in trial $t - n$, and finally, α ($0 \leq \alpha \leq 1$) is the learning rate parameter, which determines how much the prediction error (i.e., difference between perceptions of intergroup empathic norm *i* and the corresponding empathic feedback) influences updating of intergroup empathic norm perceptions (Rescorla & Wagner, 1972). One detail is particularly important to note. Specifically, in the empathy learning task, participants went through two types of trials: (a) trials where an ingroup member was an active player, and thus an empathic target, and (b) trials where an outgroup member was an active player, and thus an empathic target. On trials where an ingroup member was an active player, participants would receive empathic feedback about *ingroup intragroup empathic norms* and *outgroup intergroup empathic norms*, and on trials where an outgroup member was an active player, participants would receive empathic feedback about *outgroup intragroup empathic norms* and *ingroup intergroup empathic norms*. Given that the order of active players (i.e., empathic targets) was random, participants did not receive feedback about the same empathic norm in a consistent manner. For instance, one participant could see the following order of active players: outgroup – ingroup – ingroup – outgroup, which meant that they received feedback about ingroup intergroup empathic norms on trials 1 and 4. Another participant could instead see the following order of active players: outgroup – outgroup – ingroup – ingroup, which meant that they received feedback about ingroup intergroup

empathic norms on trials 1 and 2. For this reason, equation 1 does not include $t - 1$, but instead $t - n$, where n refers to the gap between two adjacent occurrences of the same empathic feedback.

In the second step, the Q-values (i.e., estimated perceptions of intergroup empathic norms) were used to predict participants' intergroup empathy using a linear regression formula:

$$y'^{t} = b_0 + b_1 * Q_{IN\text{-}OUT}{}^{t} + b_2 * Q_{OUT\text{-}IN}{}^{t} \quad [2]$$

In equation 2, y'^{t} refers to intergroup empathy a specific participant was predicted to show on trial t, $Q_{IN\text{-}OUT}{}^{t}$ refers to the estimated perception of ingroup intergroup empathic norms this participant had on trial t, while $Q_{OUT\text{-}IN}{}^{t}$ refers to the estimated perception of outgroup intergroup empathic norms this participant had on trial t. Parameters b_0, b_1, and b_2 refer to a participant-specific intercept, regression slope for ingroup intergroup empathic norms and regression slope for outgroup intergroup empathic norms respectively.

As outlined under hypothesis 4, I considered three main mechanisms that could have given rise to erroneous perceptions of intergroup empathic norms: (a) negative priors about intergroup empathy, (b) negativity bias wherein participants update perceptions of intergroup empathic norms more rapidly from negative prediction errors (i.e., situations where ingroup or outgroup members display less intergroup empathy than expected) as opposed to positive prediction errors (i.e., situations where ingroup or outgroup members display more intergroup empathy than expected), and (c) reciprocity bias wherein participants generalise intergroup emotional reactions of one group to the other. The six fitted RL models are outlined in Table A4.1.1 and more details on each of the models are provided following the table.

Table A4.1.1

Overview of the fitted models

Model number	Number of parameters	List of parameters
1	5	Q, α, b_0, b_1, b_2
2	6	$Q_{IN\text{-}OUT}, Q_{OUT\text{-}IN}, \alpha, b_0, b_1, b_2$
3	6	$Q, \alpha^+, \alpha^-, b_0, b_1, b_2$
4	6	$Q, \alpha, G, b_0, b_1, b_2$
5	7	$Q, \alpha^+, \alpha^-, G, b_0, b_1, b_2$
6	8	$Q, \alpha^+, \alpha^-, G^+, G^-, b_0, b_1, b_2$

Notes. Details about each model are provided in the main text under the table. While results of accuracy analyses seem to suggest that perceptions of intragroup empathic norms similarly display signs of biases in observational learning, the models outlined in this table focused only on ingroup and outgroup intergroup empathic norms. The reason for this is that these two intergroup empathic norms were of primary interest in Experiment 7 and were the only ones that were systematically manipulated.

Model 1 estimated five parameters: (a) each participants' prior about how much intergroup empathy others would display in the task (Q, $-5 \leq Q \leq 5$), (b) learning rate determining how rapidly each participant updated perceptions of empathic norms based on all available information throughout the task (α, $0 \leq \alpha \leq 1$), (c) intercept determining how much intergroup empathy each participant showed when both intergroup empathic norms equalled 0, indicating bidirectional intergroup apathy (b_0), (d) regression slope for ingroup intergroup empathic norms (b_1) and (e) regression slope for outgroup intergroup empathic norms (b_2). Importantly, Model 1 assumed that participants start the empathy learning task with varying pre-existing beliefs about intergroup empathy but then learn from all emotional reactions they observe in an identical manner. Furthermore, this model estimated a single prior, reflecting the assumption that participants have identical pre-existing beliefs about how much empathy ingroup members would display towards outgroup members and vice versa. Finally, before proceeding, it is important to emphasise that all fitted models differed only in how they

specified equation 1, while there were no differences between fitted models in how they specified equation 2. In other words, all six RL models identically estimated the intercept and the two regression slopes. Thus, to avoid repetition, these three parameters will not be further discussed under alternative models (Models 2-6).

Model 2 expanded Model 1 by allowing each participant to have different priors about how much intergroup empathy ingroup members would display towards the outgroup ($Q_{IN\text{-}OUT}$, $-5 \leq Q_{IN\text{-}OUT} \leq 5$) and how much intergroup empathy outgroup members would display towards the ingroup ($Q_{OUT\text{-}IN}$, $-5 \leq Q_{OUT\text{-}IN} \leq 5$). Based on the results of Experiment 6a reported in Chapter 3, it is unlikely that participants hold different pre-existing beliefs for ingroup and outgroup members in the context of minimal groups, but I decided to test this possibility regardless.

Models 1 and 2 incorporated different biased priors in the absence of biased updating. Model 3, however, incorporated biased updating such that participants could update perceptions of intergroup empathic norms differently (i.e., more, or less rapidly) based on negative and positive prediction errors. It achieved this by estimating separate learning rates depending on the sign of the prediction error (α^{+}, α^{-}). Finally, following Model 1, Model 3 allowed participants to have varying priors about intergroup empathic norms, but assumed those priors are the same across ingroup and outgroup intergroup empathic norms (Q).

Up to this point, in Models 1-3, participants were only allowed to learn about ingroup and outgroup intergroup empathic norms from their respective empathic feedback. In other words, participants only updated their perceptions of ingroup intergroup empathic norms when they saw ingroup members' emotional reactions to outgroup members' outcomes and only updated their perceptions of outgroup intergroup empathic norms when they saw outgroup members' emotional reactions to ingroup members' outcomes. In Models 1-3, no updating of either intergroup empathic norm took place when participants were exposed to

the other intergroup empathic norm. However, going back to the results from the accuracy analyses, in particular those from incongruent experimental conditions, one can notice that participants' perceptions of one intergroup empathic norm varied depending on the other intergroup empathic norm. For instance, in the conform-to-empathise condition, outgroup members' intergroup empathy bias was underestimated, potentially because ingroup members did not reciprocate intergroup empathy bias, while ingroup members' intergroup empathy bias was overestimated, potentially because outgroup members showed consistent intergroup empathy bias despite ingroup members showing none. These results suggest that it is likely that participants generalised one group's emotional reactions to the other group. To mimic this cross-norm learning, Model 4 introduced a generalisation parameter corresponding to the strength of each participant's reciprocity bias. In particular, Model 4 specified the following learning equations (for brevity, the equations represent only the updating of ingroup intergroup empathic norm perceptions):

$$Q_{IN\text{-}OUT}{}^{t} = Q_{IN\text{-}OUT}{}^{t-n} + \alpha * (EF_{IN\text{-}OUT}{}^{t\text{-}n} - Q_{IN\text{-}OUT}{}^{t-n}) \quad [3]$$

$$Q_{IN\text{-}OUT}{}^{t} = Q_{IN\text{-}OUT}{}^{t-n} + G * \alpha * (EF_{OUT\text{-}IN}{}^{t\text{-}n} - Q_{IN\text{-}OUT}{}^{t-n}) \quad [4]$$

Equation 3 reflects how perceptions of ingroup intergroup empathic norms are updated in Model 4 after a participant is exposed to an ingroup member's emotional reaction to an outgroup member's outcome. Equation 4 reflects how perceptions of ingroup intergroup empathic norms are updated in Model 4 after participant is exposed to an outgroup member's emotional reaction to an ingroup member's outcome. In equation 4, one can see the newly introduced generalisation parameter (G, $0 \leq G \leq 1$), which specifies how much participants learn about one group's intergroup empathic norms from the other group's intergroup empathic feedback. When G is set to 1, ingroup intergroup empathic norms are updated equally from both ingroup and outgroup members' intergroup empathic feedback, and when G is set to 0, ingroup intergroup empathic norms are only updated from ingroup members'

but not from outgroup members' intergroup empathic feedback. Importantly, Model 4 where *G* is equal to 0 corresponds to Model 1.

Model 5 combines Model 3 and Model 4 in allowing both (a) differential updating based on positive (α^{+}) and negative prediction errors (α^{-}), and (b) cross-norm learning (*G*). Finally, Model 6 is the most complex model wherein the generalisation parameter can similarly vary depending on the sign of the prediction error (G^{+}, G^{-}). Under Model 6, a hypothetical participant could, for example, update their perceptions of intergroup empathic norms more rapidly from positive than negative prediction errors ($\alpha^{+} > \alpha^{-}$) but expect greater reciprocity from the second group when the first group displays lower levels of intergroup empathy than expected (G^{-}) than when the first group displays higher levels of intergroup empathy than expected (G^{+}).

APPENDIX 4.2: Model recovery

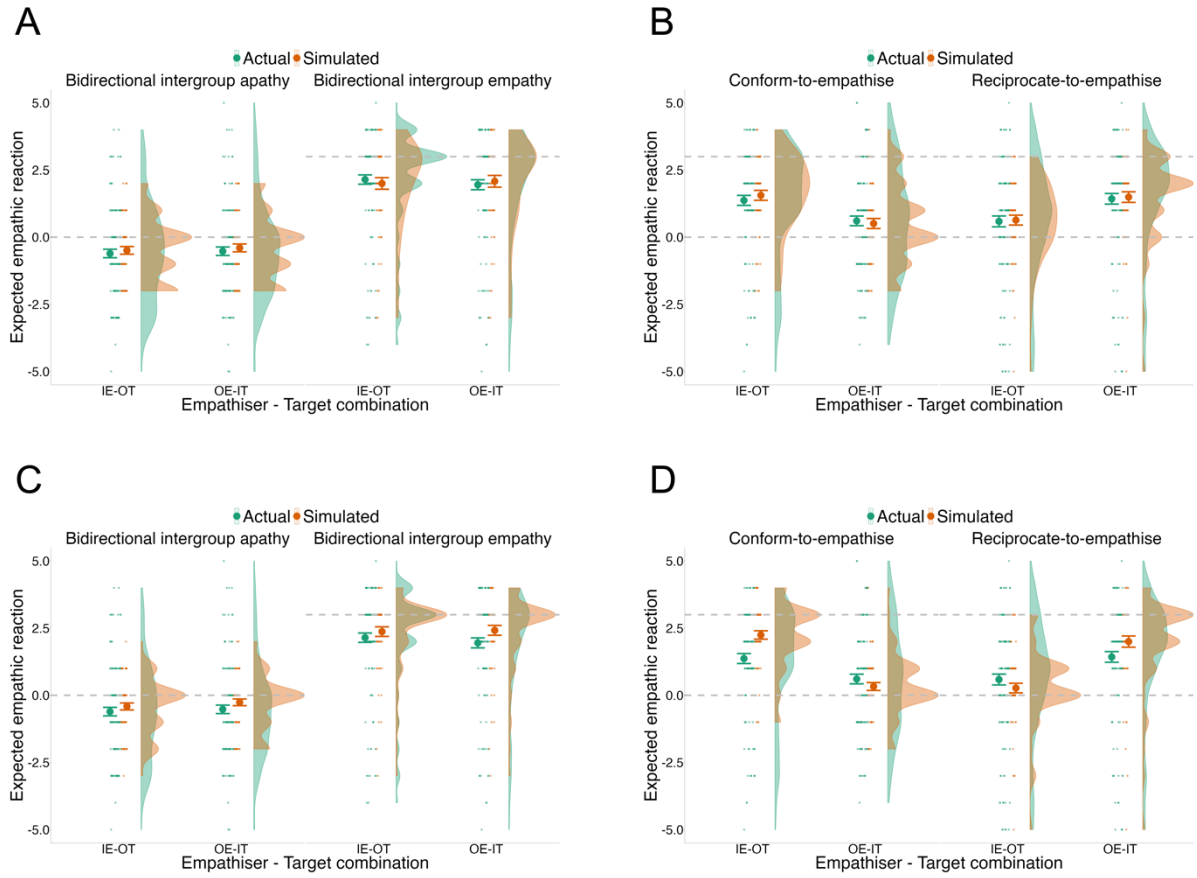
In the first step of the model recovery procedure, I obtained parameters for each participant using maximum likelihood estimation (MLE, Stillerman et al., 2020). Specifically, I used the Nelder-Mead algorithm to minimise the negative logarithm of the probability density function (i.e., negative log-likelihood) and therefore obtain parameters that maximise the joint probability density of observed data in the empathy learning task given each model. To avoid the issue of local minima, I initiated optimisation with 40 randomly selected start values for each parameter (Lockwood & Klein-Flügge, 2021). Following this same set of general steps, I took a twofold approach to parameter estimation. Under the first approach, I did not apply any priors to regularise model parameters but instead allowed the algorithm to estimate these parameters within the full range of their possible values without any penalties (i.e., learning rates and generalisation parameters could be anywhere in the range from 0 to 1, pre-existing beliefs about intergroup empathy and intercepts could be anywhere in the range from -5 to 5, while regression slopes could be anywhere between negative infinity and positive infinity). Under the second approach, I applied Beta(1.1, 1.1) priors to regularise learning rates and generalisation parameters, Norm(0, 2) priors to regularise pre-existing beliefs about intergroup empathy and intercepts, and finally, Norm (0.7, 7) priors to regularise regression slopes. Beta priors for learning rates and generalisation parameters were chosen based on prior work (Decker et al., 2015; Hackel et al., 2015) while normal priors for pre-existing beliefs, intercepts, and regressions slopes were chosen to penalise unrealistically low or high parameter values while not exerting much influence in the range of realistic parameter values. In the second step of this model recovery procedure, I simulated data under all six models (see Table A4.1.1) with the range of simulation parameters matching the range of parameter values obtained in the first (i.e., parameter estimation) step (Wilson & Collins, 2019). In the third step of this model recovery

procedure, I fitted the simulated data with all six models to determine how often the data were simulated on the basis of one model best fit by that model as opposed to any other model from Table A4.1.1. To identify the best fitting model for the data simulated from a known model, I used both the Akaike Information Criterion (AIC, Cavanaugh & Neath, 2019) and the Bayesian Information Criterion (BIC, Neath & Cavanaugh, 2012). The outcomes of this model recovery can be found in the Figure 4.7 in the main text.

APPENDIX 4.3: Simulation of intergroup empathic norm perceptions

Figure A4.3.1

Actual versus simulated perceptions of intergroup empathic norms



Notes. (A) Actual and simulated perceptions (under Model 3) of intergroup empathic norms in congruent experimental conditions. (B) Actual and simulated perceptions (under Model 5) of intergroup empathic norms in incongruent experimental conditions. (C) Actual and simulated perceptions (under Model 1) of intergroup empathic norms in congruent experimental conditions. (D) Actual and simulated perceptions (under Model 1) of intergroup empathic norms in incongruent experimental conditions. Data simulation was done in two steps. In the first step, I used the L-BFGS-B algorithm incorporated in the R function optim to obtain parameters that minimised the squared error between each participant's reported perceptions of ingroup or outgroup intergroup empathic norms and the $Q_{IN\text{-}OUT}$ and $Q_{OUT\text{-}IN}$ values estimated after the last trial in the empathy learning task. In the second step, I used the estimated parameters (Model 1: Q, α; Model 2: $Q_{IN\text{-}OUT}$, $Q_{OUT\text{-}IN}$, α; Model 3: Q, α^{+}, α^{-}; Model 4: Q, α, G; Model 5: Q, α^{+}, α^{-}, G) in combination with the empathic feedback each participant received in the empathy learning task to recreate the two vectors of Q values. From these vectors I then extracted the final $Q_{IN\text{-}OUT}$ and $Q_{OUT\text{-}IN}$ values which corresponded to each participant's final perception of ingroup and outgroup intergroup empathic norms respectively. In the upper panel (A & B), you can find perceptions of intergroup empathic norms simulated under the simplest model which could reproduce the

patterns of underestimation and overestimation observed in participants' responses, while in the lower panel (C & D), you can find perceptions of intergroup empathic norms simulated under Model 1. As can be seen in the lower panel, perceptions of intergroup empathic norms simulated under Model 1 tended to be more accurate than participants' actual perceptions. In summary, out of six considered models (Models 1-6 in the main text), Models 3 and 5 were the simplest models that could accurately reproduce patterns of underestimation and overestimation in participants' perceptions of intergroup empathic norms. However, given that these simulations do not constitute formal model comparison and that the list of potential learning models has not been exhausted, these results should be treated with caution. Gray dashed lines represent accurate levels of intergroup empathy in apathetic and empathetic conditions. IE-OT = Ingroup empathiser, outgroup target. OE-IT = Outgroup empathiser, ingroup target.