UNIVERSITY OF BELGRADE

FACULTY OF PHILOSOPHY

Marija V. Čolić

# Intergroup Empathy – Conceptual Systematisation and Empirical Contribution to Understanding the Phenomenon

Doctoral dissertation

Belgrade, 2023.

УНИВЕРЗИТЕТ У БЕОГРАДУ

ФИЛОЗОФСКИ ФАКУЛТЕТ

Марија В. Чолић

# Међугрупна емпатија – концептуална систематизација и емпиријски допринос разумевању феномена

Докторска дисертација

Београд, 2023.

Supervisor:

Full Professor Iris Žeželj, PhD
University of Belgrade, Faculty of Philosophy, Department of Psychology

Board members:

Associate Professor Marija Branković, PhD
Singidunum University, Faculty for Media and Communications, Department of Psychology

Full Professor Zoran Pavlović, PhD
University of Belgrade, Faculty of Philosophy, Department of Psychology

Full Professor Goran Knežević, PhD
University of Belgrade, Faculty of Philosophy, Department of Psychology

Date:

_____

**Acknowledgements**

First and foremost, I am grateful to my supervisor, Iris Žeželj, for all the guidance, support, and freedom she has been giving me all these years. Iris, you are the epitome of a mentor, a role model, and an inspiration.

I am also grateful to board members Zoran Pavlović, and Goran Knežević and especially Marija Branković, for their valuable questions and insights in the early stages of my work that helped me think it through.

I owe my deepest gratitude to all the people who volunteered their time and effort to help me design experimental stimuli and recruit participants: members of Brazilian Jiu-jitsu Club VIN, Alpine Club Avala, and Laboratory for Research of Individual Differences, Nikola Čolić, Vladimir Milošević, Nemanja Đorđijoski, Blagoja Radić, Vladimir Bobić, and Miloš Šaranović. I would also like to thank numerous friends, acquaintances, and strangers (especially forum administrators) who assisted with participant recruitment. To all of you: I am certain you could have spent your time more enjoyably but chose to help. Thank you!

I am especially grateful to several individuals:

Matija Ilić, my partner, besides his acting and modeling services, for not beating a hasty retreat after being obliged to listen to me dwell upon my thesis and spill my frustration over and over, endlessly. Matija, thank you for being so beautiful inside and out.

Vanja Radić, fellow BJJ practitioner and a friend, for providing filming equipment and valuable technical advice, editing the photographs, and showing a deep interest in my work and an unprecedented meticulousness in his. Vanja, your help was essential, and your attitude invigorating. Thank you.

Darko Stojilović, a colleague and a friend, for assisting with stimuli and providing valuable comments on the thesis draft, as well as for his continuous encouragement and inspiring conversations.

Dušan Mandić, a colleague and a friend, for being such a polymath, a great conversation partner, and a safe container.

I would also like to thank all the students and numerous internet volunteers who participated in the studies. This thesis would simply not be possible without you.

Finally, I would like to thank my family for being there for me and supporting me in whatever I choose to do, always. This thesis is for you.

Intergroup Empathy – Conceptual Systematisation and Empirical Contribution to Understanding the Phenomenon


Abstract

Intergroup empathy bias is a tendency to empathise more strongly with the members of own group compared to the members of the outgroup. In literature, empathy is used as an umbrella term for diverse responses to the experiences of others (from sensorimotor resonance to compassion) that implicate fundamentally different processes. The largest body of evidence for intergroup empathy bias, however, stems from neuroscientific studies. Due to the specificity of the investigated phenomenon, these studies are limited in complexity and ecological validity of study design. They also tend to disregard behavioural indicators of empathy, which limits their generalisability.

We conducted a systematic review to analyse how intergroup empathy bias is presently conceived, measured, and interpreted. We expectedly identified considerable conceptual heterogeneity; however, the results confirmed that the concept of intergroup empathy bias is primarily based on neuroscientific studies of empathic responses to physically painful stimuli. Whilst bias in neural responses is universally observed, it is not the case with the behavioral responses. This discrepancy was rarely addressed and ascribed to social desirability We argued that applying these findings to the entire concept of empathy is premature and could thus be misleading.

Informed by the systematic review, we designed four experiments focusing on empathic responses to physical pain. To exclude the social desirability explanation, we opted to study football fan groups. While the first two experiments were conceptual replications of frequent empathy-eliciting paradigms in neuroscience research on a substantially larger sample, in the latter two we designed more complex painful stimuli to address the effects of ecological validity and contextual embeddedness of painful events on a broader range of empathic responses.

We observed no bias when participants were asked for decontextualised pain assessment. When pain was assessed with complex stimuli, we observed bias only when the context was directly related to the participant's group identity.

This pattern of results fitted well in the expectations derived from the social identity theory: bias in empathic responses to painful events emerged only when the social identity was accessible and provided an appropriate framework for responding to a social situation. The bias was shaped by individual (identification) and social variables (group-based threat). We observed both ingroup favouritism and outgroup derogation in different empathic responses.

Summary outcomes of the review and experimental studies were discussed in the context of meta-scientific issues in empathy research and research of bias



Keywords: intergroup empathy bias, empathy for pain, motivated empathy, social identity theory, group identity, motivated social cognition, meta-science

Scientific field: Psychology

Subfield: Social psychology

UDC number:

Међугрупна емпатија – концептуална систематизација и емпиријски допринос разумевању феномена

Апстракт

Међугрупна пристрасност у емпатији је склоност да имамо више емпатије према припадницима група којима припадамо у односу на групе којима не припадамо. У литератури емпатија се користи као „кишобран“ термин за разноврсне одговоре на искуства других (од сензомоторне резонанце до саосећања), који подразумевају фундаментално различите процесе. Међутим, највише доказа за постојање ове пристрасности потиче из неуронаучних студија. С обзиром на специфичност испитиваног феномена, ове студије су ограничене по питању комплексности и еколошке валидности дизајна. Додатно, понашајни индикатори емпатије су често занемарени, што ограничава могућност генерализације налаза.

Како бисмо анализирали како се међугрупна пристрасност у емпатији тренутно дефинише, мери и тумачи, спровели смо прегледну студију. Очекивано, идентификовали смо значајну хетерогеност у појмовима; међутим, резултати су потврдили да је концепт међугрупне пристрасности у емпатији примарно заснован на неуронаучним студијама емпатије за физички бол. Иако је пристрасност у неуралним одговорима забележена у свим студијама, то није случај са понашајним. О овој несагласности се ретко дискутује и приписује се социјалној пожељности. Указали смо зашто сматрамо да је примена ових налаза на читав концепт емпатије преурањена и како може навести на погрешне закључке.

На основу прегледне студије осмислили смо четири експеримента, са фокусом на емпатијским одговорима на физички бол. Како бисмо искључили социјалну пожељност као објашњење, проучавали смо навијачке групе. Прва два експеримента су представљала концептуалну репликацију често коришћених парадигми за изазивање емпатије у неуронаукама на значајно већем узорку испитаника, док смо за друга два експеримента осмислили комплексне болне стимулусе како бисмо испитали ефекте еколошке валидности и контекстуализације болних догађаја на ширем скупу емпатијских одговора.

Међугрупна пристрасност није забележена када смо од испитаника тражили деконтекстуализовану процену бола. Када је болна стимулација илустрована комплексним стимулусима, забележили смо пристрасне емпатијске одговоре само када је контекст био директно релевантан за групни идентитет испитаника.

Овај склоп налаза добро се уклапа у очекивања теорије социјаног идентитета: пристрасност у емпатијским одговорима на болне стимулусе јавила се тек онда када је групни идентитет био доступан свести и када је пружао адекватан оквир за одговарање на социјалну ситуацију. Оба облика пристрасности – фаворизовање своје групе и дерогирање друге групе – забележена су у различитим емпатијским одговорима.

Збирни исходи прегледне и експерименталних студија дискутовани су у контексту мета-научних питања у истраживању емпатије и истраживању пристрасности.

Кључне речи: међугрупна пристрасност у емпатији, емпатија за бол, мотивисана емпатија, теорија социјалног идентитета, групни идентитет, мотивисана социјална когниција, мета-наука

Научна област: Психологија

Ужа научна област: Социјална психологија

УДК број:

# Table of Contents

## Defining and Measuring Empathy

*"So you think, do you, it is only houses that are built? I am continually building myself and building you, and you are doing the same, inversely."*

— Luigi Pirandello, Uno, nessuno, e centomila (1926)

Pirandello's protagonist (Pirandello, 1926/2007), Vitangelo Moscarda, becomes consumed by a metaphysical identity crisis after one morning his wife points out that his nose leans slightly to the right. Having never noticed this himself, he realises that his own image of himself and his wife's Genge are not the same person. Applied universally, this led to a conclusion of an inevitable mutual misapprehension, as everyone sees themselves in certain ways not corresponding to how anybody else sees them and sees others in myriad different ways not corresponding to their self-image. Vitangelo becomes obsessed with observing everything through an identity-related philosophical lens and his everyday behaviour gradually turns to be so detached from reality that his relationship with his family and the community deteriorates. This all-consuming identity crisis ultimately led to self-destruction.

Empathy, like Vitangelo, is suffering an all-consuming identity crisis, resulting in an inevitable mutual misapprehension. It is indeed a paradox that a concept used to explain how we can understand the experiences of others is plagued with issues in self-understanding. Everyone has their own "empathy" – it is one of those infamous constructs in psychological science that has as many definitions as the number of authors studying it. Batson (Batson, 2009) lists as many as eight different concepts labelled empathy in contemporary research practice: to know the inner state of another, to resonate with the inner state of another, to feel what another feels, to put yourself in place of another, to imagine what another person thinks or feels, to imagine how you would feel at their place, to be upset by the sight of another person's suffering, and to feel with another person. Intuitively, these concepts seem related; however, the psychological processes they implicate can be fundamentally different. In neuroscience, empathy is regarded as a multidimensional construct consisting of separable but related components working in parallel: emotion/experience sharing, perspective taking, and empathic concern (Decety & Cowell, 2014), suggested to depend on separate brain structures.

If we consider all empirical studies in which the research subject is labelled or subsumed under "empathy" regardless of the method, it is evident that both of the presented divisions include some empathic components under empathy that are explicitly excluded in other definitions, for example, empathic distress (Preston & De Waal, 2002; Cuff, Brown, Taylor, & Howat, 2014). In addition, empathy is typically not defined but considered self-explanatory in an enormous number of studies (especially applied studies). On the other hand, studies that do define empathy explicitly often define it broadly but measure a substantially more narrow or specific concept (Hall & Schwartz, 2018).

There have been attempts to make a conceptual systematization in empathy research, that targeted definitional issues. In a recent analysis of the use of the term whose aim was to provide an all-encompassing definition, as many as 43 different definitions and eight different points of dispute were identified: how to differentiate empathy from related concepts, whether is affective or cognitive, if it is possible without direct experience, is it a state or a trait, automatic or controlled, etc). The authors discuss point by point, providing arguments for their view in each, thus building a summary definition of empathy that at the same time precisely excludes all those phenomena that should not be considered empathy. The resulting definition (Cuff et al., 2014) states that: *"Empathy is an emotional response (affective), dependent upon the interaction between trait capacities and*

*state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with a recognition that the source of the emotion is not one's own."* (p. 150).

However, some authors are against these summary definitions consisting of long lists of conditions defining the phenomenon (and the non-phenomenon), believing they only contribute to the conceptual confusion; they advocate creating a taxonomy of empathy-related phenomena or abandoning the general term in favour of more specific concepts (Decety & Cowell, 2014; Hall & Schwartz, 2018; Zaki & Cikara, 2015). In our view, a taxonomical classification of empathic phenomena with regards to similarities and differences in their definitions and measurement simultaneously would greatly contribute to our understanding of the phenomena and their reach in predicting real behaviour. In this thesis, we are going to focus on empathy measurement.

The final answer to the question *What is empathy?* is beyond the scope of this thesis. However, we believe it is important for our research problem to highlight the diversity of the phenomena being studied under the same name because of the importance of the concept of empathy in both theoretical and applied terms. Namely, by studying various emphatic phenomena from motor mimicry to imaginative identification with movie characters we are trying to answer two broad questions: how do we get to know what another person thinks or feels, and what motivates us to feel concern for another person and help them in trouble. The former is the fundamental question of the emergence and functioning of human social life (Preston & DeWaal, 2002), or the question of abilities, states, traits, capacities, responses, and behaviours (Hall & Schwartz, 2018) that enable us to recognise and know the internal states of another person, coordinate our actions and live within a community. The latter is the question of how, if, when, and under which conditions could that feeling be manipulated to minimise or prevent negative social consequences, or, in other words, this is the question of empathy intervention studies.

For both questions, it is important to clearly determine what we are talking about when we say empathy. It is inherently important because by being conceptually precise we are facilitating the systematization, integration, and advancement of our knowledge about the phenomena, which enables us to make more precise practical implications and recommendations. However, conceptual precision is also essential for the praxis of scientific communication between (sub)disciplines. Review articles, which are most relevant for interdisciplinary communication because they convey most information with as few technical details as possible, put very different phenomena under the umbrella term "empathy", which most often ends up in its non-differentiated form in a loosely connected field of study as a theoretical rationale. Precisely defining the phenomenon under study and its relationship to related phenomena would facilitate our reasoning about how and to what extent is a specific research finding relevant for designing a study or an intervention in which we manipulate empathy to increase prosocial and reduce antisocial behaviour.

In addition, conceptual clarity in empathy research is also important outside of scientific and terminological disputes, as the construct is widely used in public discourse and lay explanations of human behaviour in everyday communication. In fact, some authors ascribe the recent popularization of the word empathy to the discovery of so-called mirror neurons as a hypothetical neural substrate of intersubjectivity and believe this only contributed to the conceptual disorder (Milivojević, 2015). Therefore, if we clearly delimited the empathic phenomenon under study and placed it in a system of human responses to others' states, we would contribute to a better understanding of its explanatory power for everyday behaviours that raise curiosity and concern in the lay public and facilitate precise communication of scientific knowledge.

The phenomenon we were studying in this thesis, intergroup empathy bias, is affected by all of the issues we raised. Under intergroup empathy bias we mean differential empathic responses towards members of our own group compared to another group, ranging from decreased sensorimotor resonance to pain of the outgroup (e.g. Avenanti, Sirigu & Aglioti, 2010) to

difficulties in understanding their mental states (e.g. Adams et al., 2010). Obviously, very different empathic responses were being studied under the label of intergroup empathy bias. Contrary to present research practice, we believe that putting them all under the umbrella term "empathy" and treating them as manifested forms of unitary psychological processes hinders our knowledge by oversimplifying it and may hinder our motivation to look further by creating an impression that we know about it much more than we currently do. Therefore, we are going to approach the intergroup empathy bias study with variation in mind, with a particular focus on empathy-eliciting tasks and measurement.

## Intergroup Empathy Bias – Mapping the Concept

It is hard to overestimate the importance of studying empathy in intergroup relations. On the one hand, empathy and empathic constructs are intuitively, conceptually, and empirically positively related to prosocial behaviour and negatively related to antisocial behaviour of different kinds. Moreover, it has been demonstrated that prosocial behaviour can be encouraged and aggressive behaviour reduced with interventions aiming to increase empathy, not only in interactions between individuals but in intergroup contexts as well (Eisenberg, Eggum, & Di Giunta, 2010). On the other hand, there is a growing number of studies pointing out that empathy has its limits and they are often drawn based on group affiliation (Cikara, Bruneau, & Saxe, 2011), inviting us to carefully analyse and consider the idea that more empathy is always better.

Intergroup relations are more competitive than interindividual interactions in general (De Dreu, 2010), and empathy "failures" are considered to have an important role in many examples of social conflicts, from hate crimes to large-scale armed conflicts between national or religious groups (Cikara, 2018). Moreover, some authors such as Paul Bloom (Bloom, 2016) are explicitly "against" empathy, believing that empathy is parochial or biased by definition and that its emotional nature and its boundedness to what is in focus (and that is most frequently one person or a small number of individuals) hinder our ability to make rational and moral decisions. Although we do not share the opinion that empathy is "morally corrosive", we believe that a careful analysis of these "failures" and parochial qualities would contribute to our understanding of the nature of empathy, but also to finding ways to design, conduct and evaluate interventions aimed at decreasing negative social phenomena such as prejudice, discrimination, and violence.

According to the review studies of intergroup empathy bias (e.g. Cikara, Bruneau, & Saxe, 2011; Fourie, Subramoney, & Madikizela, 2017), it seems that this problem has mainly been studied in neuroscience. The nature of the data and measurement in these studies do not allow for conscious and willful control over the reactions being measured, which provides an important advantage of neuroscientific studies compared to self-report and other behavioural measures of empathy more frequently used in social psychology (Hall & Schwartz, 2018; Neumann, Chan, Boyle, Wang, & Westbury, 2014). The issue of social desirability is present in both empathy research and research on ingroup bias, and especially so in research on intergroup bias in empathy, because, it is socially expected to be both empathic and unbiased. By using brain imaging techniques this issue is circumvented. Moreover, neuroimaging and neurostimulation techniques provide insight into processes difficult to measure in another way because they are automatic and impossible to consciously assess, and too subtle for crude behavioural measures such as accuracy and response time. Besides, these techniques can provide direct insight into how processes leading to intergroup bias emerge and develop (Scheepers & Derks, 2016) (because behavioural measures record only the outcome of the process, and the process itself is being inferred). Finally, the results of neuroscience studies are valuable on their own as an additional source of information about a phenomenon and could provide empirical evidence for one of the alternative explanations of the

phenomenon that have been previously laid out based on the theoretical and empirical corpus (Tingley, 2006).

However, neuroscience studies are limited with respect to the complexity of the stimuli that can be presented and reliably measured, or in other words, with respect to their ecological validity. Consequently, the methodology of neuroscientific research restricts the range of phenomena in the domain of empathy that could be studied. Despite this important limitation, the results of neuroscience studies are frequently discussed as a contribution to our knowledge about empathy as an umbrella term, and hypotheses are being made about the implications of empathy bias registered in these studies for constructs such as intergroup violence (Cikara, 2015) or helping behaviour in intergroup context (Eisenberg et al., 2010). In addition, results of neuroscience studies are used as an argument in fields of study that are very remotely connected to the empathic phenomenon being measured (for example in law studies in discussions for and against capital punishment, Johnson, Hritz, Royer, & Blume, 2016), but also in designing clever and technologically advance interventions (for example, Bertrand, Guegan, Robieux, McCall, & Zenasni, 2018).

To illustrate, we will take as an example the result of a study (Avenanti et al., 2010) that people resonate less with the pain of a person of a different race compared to people of their own race when they see a needle on a computer screen allegedly penetrating black or white hand. This result is being interpreted as a suspension of empathy and disinhibition of aggressive behaviour in intergroup conflicts, or even further, based on this finding, it is being suggested that we could use virtual body illusions (simulation of presence in another body) to decrease or eliminate this intergroup empathy bias in general. Although this characterization of research logic may be too simplified and harsh, it points to the need to clarify and specify what we have actually studied and how far we can spread our conclusions. Besides, with the impact of studies investigating empathic phenomena and the sheer number of interventions that target empathy in mind, we believe that it is necessary to make an effort to investigate intergroup empathy bias with stimuli closer to the everyday laypeople experience, to ensure that our knowledge could be used in practice more directly and unequivocally.

Our preliminary literature search on empathy and intergroup bias resulted in a number of tentative conclusions. Most behavioural studies speaking of empathy in intergroup contexts investigate the relationship of empathic responses with dispositional measures related to discrimination or with designing interventions that target empathy – in other words, they do not directly measure the differential empathic response to ingroups and outgroups. Among the studies that do, the most prominent place and substantial weight are given to studies registering differential neural responses to physical pain suffered by ingroups compared to outgroups. In other words, it seems that for intergroup empathy bias, our conclusions about empathy, in general, are mostly based on one of the countless hypothetical quadrants – neuroscientific studies of empathy for physical pain. As resonating with the internal state of another is one of many empathic reactions (and here we will skip problematizing the equality sign between the concept and the actual measure of neural activity), generalising biased neural responses to physically painful stimuli to the entire concept of empathy is unjustified and may be misleading.

However, physical pain is especially interesting as it represents a basic and universal human experience and is as such hypothesised to elicit a universal empathic reaction in the observers. Therefore, to inflict pain on another person, one needs to suspend, control, modify, or fail to develop an empathic reaction to the signals of pain another person is emitting. Inflicting pain on other people is considered to be morally wrong in almost every culture, and people who enjoy it are seen as deviant and their behaviour is sanctioned both formally and informally by the community. However, in intergroup contexts, inflicting pain on other people can be presented (to self and other people) as morally justified or even necessary (Bloom, 2016). It is important to know the exact mechanism of that switch to be able to determine how we could prevent or minimise it, and if it is possible to tackle it by interventions aimed at an individual at all.

Besides the problem of high specificity or low ecological validity of the stimuli, there is a broader issue with the manner ingroup bias is studied in neuroscience, which is sometimes acknowledged within the discipline. Namely, in spite of being inspired and supported by theoretical notions and empirical regularities discussed and interpreted within specific socio-psychological theories (social identity approach, Tajfel & Turner, 1979; realistic conflict theory, Jackson, 1993; stereotype content model, Fiske, Cuddy, Glick, & Xu, 2002), neuroscience studies of bias frequently come down to a description of differences in which structures the existing groups i.e. "us" and "them" are represented in the brain (Cikara & Van Bavel, 2014). In other words, the majority of studies investigate if in certain parts of the brain previously shown to be active during empathic responses the activity differs depending on the group identity of the target, and the results are being discussed in the context of neuroscientific accounts on empathy and other neuroscience studies. Integration and discussion of results with regard to important concepts in the study of intergroup relations and theoretical explanations and points of dispute in studying ingroup bias in social psychology (for example, whether the bias is achieved by derogating outgroup members or favouring ingroup members) is almost entirely absent. In addition, by using existing groups that have rich, complex, and different histories, and that can be differentiated by several criteria relevant to the emergence of ingroup bias (e.g. status differences, perceived competitiveness, specific stereotypes associated with the group, etc.), it is not possible to separate the direction and intensity of the specific effects each of these differences lead to.

Lastly, in accordance with the dominant focus on differences in brain responses to ingroup and outgroup members in neuroscience studies, the behavioural indicators of bias in these studies are not regarded highly. Under "behavioural indicators" we mean different indicators – from response times, and classification accuracies, to conscious estimates of intensity and valence of emotions displayed by the target and our own emotional responses to those displays – that capture changes in participants' visible and measurable behaviour. At the beginning of the millennia, it was necessary to argue for the importance of neuroscientific indicators as compared to behavioural in cognitive neuroscience (Henson, 2005; Wilkinson & Halligan, 2004). While emphasizing that we do not know enough about the structure of cognition to claim with high certainty what is the appropriate or expected relationship between behavioural measures and indices of brain activity, cognitive neuroscientists highlighted that brain activation represents a piece of meaningful information when it is followed by differences in behavioural indices but also when it is not followed by behavioural consequences.

Presently, it is our impression that in intergroup empathy bias studies (and social neuroscience studies of bias in general), the relative standing of behavioural versus neural indices is reversed, with neural activity being prioritised, which leads to several issues in interpretation.

a) The differences in empathic responding to ingroups and outgroups manifested in directly observable and measurable behaviour are not always assessed or shown, and if they are, they are poorly discussed or not at all. This practice points out that behavioural measures of bias are less valued.

b) Measures of behaviours that develop by relying on substantially different processes and which are subject to conscious control and sensitive to social desirability concerns to a variable degree are all grouped based on their similarities while disregarding their differences. The differences in sensitivity to social desirability concerns are especially important for interpreting the relationship between behavioural and neural indices. However, judging by the reviews, neither social psychologists nor neuroscientists investigating intergroup empathy bias in brain responses reflect on this relationship, or when they do, social desirability is evoked as a post hoc explanation for every situation when the two types of measures do not match (e.g. Amodio, 2014; Cikara & Fiske, 2011; Han, 2018) (when they do match – there is no problem, hence no need for discussion). However, participants' unwillingness to report their true feelings and assessments is one of the possible reasons why the differential response can sometimes be detected only in brain

representations of us and them, but not the only reasonable explanation. For example, it is possible that processing information about ingroups and outgroups relies on partially different processes and structures that functionally lead to the same outcome. It could also happen that the difference in brain activations is too small to be detected in behavioural responses.

c) Even if we record no differences in brain activity in two sets of stimuli, although it seems straightforward it is not necessary that those stimuli were processed in the same manner. Theoretically, information processing could rely on the same structures and levels of activity of those structures, but different processing algorithms.

d) Behavioural measures represent a validation criterion that certain differences in brain activity could be ascribed to the experimental manipulation in focus, and not to some other confounding variable related to the structure of the task or measurement procedure.

e) Finally, behavioural measures describe the phenomenon on a level that cannot be reduced to brain representations, and that could be equally important for understanding the phenomenon and predicting its relationship to theoretically and practically relevant concepts. In other words, if our goal is to predict and control human behaviour in the real world, an effect that cannot be related to any measurable behaviour may be completely useless for that purpose.

As we have already discussed, our knowledge about the relationship between what happens in the brain compared to what manifests in behaviour is still rudimentary to draw conclusions about psychological processes relying on neural activation only, even for basic cognitive processes such as form perception (Wilkinson & Halligan, 2004), and especially if we are speaking about complex behaviours that unfold and are defined by social context (Tingley, 2006). The question of how the knowledge we derive from brain activation can contribute to our knowledge about realistic behaviours such as voting behaviour has been raised in some fields of study, for example, political science. One of the views on neuroscience studies we also endorse is that these data are useful to the extent they can contribute to generating specific and testable hypotheses that rely on and critically review our current knowledge about a certain social phenomenon. In addition, it is emphasised that the search for new data sources must go in parallel with the conceptual development and must include a critical discussion about the theoretical and empirical limitations of all sources of data, including brain activation (Tingley, 2006). To our knowledge, in social psychology, such a focused discussion is currently lacking. Distinguished authors in the neuroscience of intergroup relations highlight the need for theoretical integration of divergent studies in the field and for incorporating familiar and robust socio-psychological concepts in both the design and discussion of studies measuring brain activity (Cikara & Van Bavel, 2014). To the best of our knowledge, there are no studies answering that call, therefore our intention is to start contributing by filling in the gaps. In addition, our opinion is that both agreement and disagreement on two levels of measurement can inform theories about the phenomenon we are studying – both socio-psychological and neuroscientific – and contribute to a better understanding of human behaviour and the way it unwinds in everyday life.

In summary, based on review articles it seems that the majority of intergroup empathy bias studies are neuroscience studies and that they dominantly investigate empathy for physical pain suffered by a member of your own compared to another (most frequently racial) group. Results are discussed primarily in neuroscientific terms and with regard to neuroscience models of empathy, and the integration with socio-psychological explanations of bias is missing. Consequently, behavioural indices of empathic responding, although they are a step closer to predicting and controlling behaviour, are neglected in interpretation.

We will now present a review of theories and concepts from social psychology important for studying intergroup empathy bias, followed by a discussion of specific intergroup empathy bias studies with respect to those concepts. As empathy is something, everything, and nothing at all (just like the poor Vitangelo), we will limit our presentation (and this thesis) to articles that (1) treat

empathy as a dependent variable with respect to group identity and (2) explicitly label the response they measure an empathic reaction.

## Systematisation and Problematisation of Empirical Findings

To truly grasp to what extent we are connected to groups we identify with and how strongly they define us as individuals, try answering the question *Who am I?* without referring to any of them. Many of us will find it difficult to list more than a few attributes without mentioning any group affiliation. A part of our identity derived from our knowledge that we belong to a certain group together with the value and emotional significance we ascribe to that group is called social identity (Tajfel & Turner, 2004). Strongly identifying with a group facilitates social categorization and amplifies its psychological consequences: people not only chronically think about themselves as part of the group, but think, feel, and act on behalf of the group (Mackie, Smith, & Ray, 2008; Smith & Mackie, 2008). When our social identity is salient, classifying oneself as a group member leads to a switch from individual to collective self-concept, which has far-reaching consequences on the way we perceive, evaluate, and react to members of other groups (Tajfel & Turner, 2004). In fact, reviews of neuroscientific studies of social categorization point that this process can be identified in the brain: during the process of categorizing ingroups and outgroups same areas of the brain are activated relatively reliably (Scheepers & Derks, 2016).

The propensity to evaluate members of groups we belong to more positively compared to non-members in domains of cognition, attitudes, and behaviours is called intergroup bias, and it can be manifested in forms of stereotypes, prejudice, and discrimination (Hewstone, Rubin, & Willis, 2002). Generalisations about members of other groups are more frequently negative or ambivalent, charged with negative emotions, and can lead to negative behaviours towards others for no reason other than their group affiliation.

Although the importance of empathy for social interaction and prosocial behaviour is frequently emphasised in empathy studies, we have already presented the view that empathy is neither universal nor sufficient for positive outcomes (Cikara, Bruneau, & Saxe, 2011). Not only that we often fail to empathise equally with socially distant targets compared to those whom we consider close, but we sometimes respond with opposite feelings such as Schadenfreude, a malicious feeling of joy following the misfortunes of other people, which according to some views can increase tolerance for or even facilitate participation in acts of collective violence (Cikara, 2015). Intergroup empathy bias manifests as differential empathic response to targets depending on their group affiliation, i.e. diminished empathic reactivity to ingroup compared to outgroup members. The existence of this bias has been demonstrated in experience sharing, perspective-taking, and emphatic concern (Eres & Molenberghs, 2013; Fourie, Subramoney, & Gobodo-Madikizela, 2017; Molenberghs, 2013).

According to the social identity approach, it could be argued that intergroup empathy bias, like other kinds of ingroup biases, is an expected consequence of group categorization and identification (Hornsey, 2008). Social identity theory (SIT) and social categorization theory (SCT) presuppose that people are generally motivated to maintain a positive self-concept, and as group identification implies incorporating group identity in our self-view, it is necessary to ensure that our own group is positively defined. In an intergroup context, this means better than other groups, and the comparative advantage can be achieved with biased perception, evaluation, and behavioural responses. Moreover, even evolutionary models that see empathy as a universal response to experiences of other people that develops early during ontogenesis and enables us to adopt social

and moral norms do not imply that in adulthood this response will be equal for everybody (Preston & De Waal, 2002).

On the other hand, intergroup empathy bias is often referred to as a "failure" of empathy, implicitly (and sometimes explicitly) implying that something "went wrong" (e.g. Zaki & Cikara, 2015). These opposite expectations are related to the question of whether intergroup bias (a difference in responses) is achieved by favouring our own groups and their positive qualities ("ingroup love") or derogating other groups by pointing out their flaws ("outgroup hate"). The former is seen as an initial form of discrimination occurring before the interaction has taken place and is often described as comprising of blind extension of trust and empathy to members of our own group, positive evaluation, and willingness to cooperate with them. That is, the fairness principle is being violated by evaluating ingroup members more positively compared to everyone else. This asymmetry in our relationship to the ingroup compared to outgroup members is achieved exclusively by favouring the ingroup and is different from outgroup derogation, which has a proactive and aggressive component (Hewstone, Rubin, & Willis, 2002). Namely, outgroup derogation implies violating the fairness principle by negatively evaluating and discriminating against outgroup members. Importantly, outgroup derogation is a negative evaluation and hostile behaviour towards a specific, concrete outgroup (as compared to or generalised "others").

In summary, according to dominant theories of intergroup bias, empathy bias is arguably an expected consequence of group division and intergroup differentiation, meaning it necessarily follows the division. This is in line with the writings of Paul Bloom (Bloom, 2016), who sees emotional empathy as biased by definition, and an interfering influence in making morally correct decisions (and thus morally corrosive).

Most socio-psychological studies of bias in general indicate that ingroup favouritism is the dominant form of bias; however, it is possible that studies on extreme groups will lead to a more frequent detection of outgroup derogation (Hewstone, Rubin, & Willis, 2002). Studying the mechanism of creating bias is important for our understanding of influences leading to one or another form of bias, as well as their possible consequences and outcomes for intergroup relations. Therefore, in further text, we will address how ingroup favouritism versus outgroup derogation was discussed in previous studies of intergroup empathy bias.

### *Intergroup Empathy Bias to Own- Compared to Other-race Members*

Intergroup empathy bias has been most extensively studied with respect to racial group membership. Racism undoubtedly was and unfortunately still is one of the most important causes of mass discrimination, collective violence, and inequality all over the world. In addition, racial categorization happens within milliseconds (Bartholow & Ito, 2009) and is based on prominent visual differences, which makes it suitable for experimental manipulation. Moreover, the disproportional number of studies of racial bias (compared to e.g. ethnicity, class, gender) in social psychology in general, can be ascribed to the sheer number of researchers from North America where racial discrimination is a prominent and widely publicly discussed social issue (see e.g. Christopher, Wendt, Marecek, & Goodman, 2014)

There are several behavioural studies on racial groups that treat empathy as a dependent variable with respect to group membership. For example, racial bias in empathy has been studied in the medical context. Pain is an aversive experience and observing others who are suffering from pain i.e. detecting verbal and non-verbal signals of pain in other people (or animals) is a strong trigger for empathic reactions and caring and soothing behaviours (Preston & De Waal, 2002). However, the analyses of medical records in the USA point out that the pain of Black individuals is

systematically underestimated and consequently treated inappropriately (Trawalter, Hoffman, & Waytz, 2012), and it is often hypothesised that empathy bias is to blame. Although theoretically, lack of empathic concern for the pain suffered by other race members could contribute to biased pain intensity estimates, experimental data from a series of studies point to both Caucasian and Black participants, medical personnel included, initially suppose that Black patients feel less pain. This was related to the participants' beliefs that hardships Black people had to endure during their lifetime made them more resistant to pain (Trawalter, Hoffman, & Waytz, 2012). These findings illustrate how concrete expectations of a specific group shape our empathic responses and not the lone fact that the person suffering pain is not a member of our group. In addition, in a series of behavioural studies, Dovidio and collaborators (Dovidio et al., 2010) demonstrated that empathy was related to prosocial behaviours in an intergroup context, but when we treat empathy as a dependent variable, it was also related to certain negative outcomes (such as empathic vengefulness towards other- but not own-race members, which could be interpreted as an aggressive act).

Racial bias in empathy has been investigated in a number of studies using neuroimaging techniques to measure empathic responses. Studies in which empathy was elicited with visual stimuli depicting physical pain (picture-based paradigms) focus on differential responses to ingroups and outgroups in the anterior cingulate cortex, anterior insula, and other components of the "pain matrix", a system of connected structures activated both during experiencing and observing pain (Lamm, Decety, & Singer, 2011). Several fMRI and EEG studies point out that our neural reaction is stronger when we observe members of our own race suffering painful stimulation and less strong for other-race members, and that information processing and the emergence of brain empathic reactions follow a different time course. (fMRI: Azevedo et al., 2013; Cao, Contreras-Huerta, McFadyen, & Cunnington, 2015; Wang, Wu, Liu, Wu, & Han, 2015; Xu, Zuo, Wang, & Han, 2009; EEG: Han, Luo, & Han, 2015; Riečanský, Paul, Kölble, Stieger, & Lamm, 2015; Sessa, Meconi, Castelli, & Dell'Aqua, 2014; Sheng, Du, & Han, 2017; Sheng, Han, & Han, 2015; Sheng & Han, 2012; TMS: Avenanti et al., 2010). We notice that bias was described as increased neural activity while observing ingroup pain and decreased neural activity while observing outgroup pain alternately and simultaneously, sometimes within one article (e.g. Azevedo et al., 2013). Indeed it is almost impossible to choose an adequate neutral group in interracial relationships, which is necessary to determine if the intergroup bias is created by favouring the racial ingroup or derogating the racial outgroup. Among the articles we listed, only three studies included some kind of control group, by presenting a purple hand being penetrated by a needle along with a black and a white one (Azevedo et al., 2013, Riečanský et al., 2015; Avenanti et al., 2010). However, a purple hand is an inadequate control situation for many reasons, but primarily because participants do not see it as human, which limits their empathic response by definition. Despite the acknowledged difficulties, it can be argued that an important theoretical distinction in socio-psychological explanations of intergroup bias is disregarded in study design.

Some of these studies tested and did not find a significant relationship between intergroup empathy bias in brain responses and any behavioural indicator of empathy – explicit assessment of pain intensity and observers' self-unpleasantness; we have previously discussed that without behavioural validation it is problematic to claim a certain effect exists (e.g. Cao et al., 2015; Xu et al., 2009).

Instead of comparing behavioural and neural empathic responses, some studies focus on the relationships between brain indicators of empathy bias and implicit and explicit behavioural measures of general ingroup bias. Neuroimaging studies of empathy are sometimes labelled "implicit empathy paradigms" (Coll, Viding, Rütgen, Silani, Lamm, Catmur, & Bird, 2017), mirroring the differences between implicit and explicit measures of prejudice. In studies of ingroup bias, the weak relationship between explicit and implicit measures of bias is sometimes interpreted as a consequence of the fact that those measures assess distinct and distinguishable constructs and processes, and that those could be differentially related to real behaviours (e.g. in public versus private contexts). It is believed that normative context, i.e. the degree to which the division between

us and them is a sensitive issue, is the most important influence in determining whether implicit and explicit measures will be related (Hewstone, Rubin, & Willis, 2002). Following this reasoning, it could be expected that implicit measures of prejudice are more directly related to intergroup empathy bias in brain responses than explicit measures of prejudice, presuming that social desirability concerns strongly influence the latter but not the former. However, the evidence for this is inconclusive. For example, some studies demonstrated that intergroup brain bias can be predicted by implicit measures of racial prejudice (namely, the IAT; e.g. Avenanti et al., 2010); however, others fail to detect a significant relationship (e.g. Riečanský et al., 2015). We have already pointed out the issues with using social desirability as a universal explanation for every discrepancy of behavioural indicators and neural indicators of bias, and the systematization of findings with respect to task demands should facilitate our understanding of influences that shape empathic responses on an implicit and explicit level.

Some of these studies tested the relationship between biased empathic brain responses and dispositional empathy (conscious and explicit trait self-assessment). The main idea was to disentangle the relationship between trait empathy and state empathy as measured by neural resonance – is ingroup bias positively, negatively, or unrelated to our general propensity to empathise with others. However, the results were not unanimous once again. In some of these studies, trait empathy predicted only own- but not other-race responses (e.g. Sessa et al., 2014), in some studies trait empathy was related to empathic reactivity in general (regardless of group identity (e.g. Sheng et al., 2017), and some studies detect no significant relationship to brain responses (e.g. Cao et al., 2015; Xu et al., 2009).

Studies that use stimuli depicting emotional pain (complex visual and verbal scenes of emotionally painful situations) point out that cultural influences can shape the nature of empathic responses in different ways for different groups. For example, participants from Korea showed stronger intergroup empathy bias in both behavioural and neural responses (greater activation of the tempo-parietal junction, a structure previously shown to be important for representing and inferring the mental states of others) compared to Caucasian participants (Cheon et al., 2011). A similar result was obtained in a study with Black participants (Mathur, Harada, Lipke, & Chiao, 2010), where the medial prefrontal cortex (mPFC) was activated in addition to other structures when participants observed the pain of an ingroup compared to an outgroup. This study was interpreted as pointing to "ingroup love" as the dominant form of bias as an entire additional brain structure was activated when observing ingroup pain. The medial prefrontal cortex is a brain structure important for social identification, and its activation depends on the strength of identification with our own group (Mathur, Harada, & Chiao, 2011; Molenberghs & Morrison, 2014), which is in line with findings from behavioural studies of bias that point to stronger group identification in minority groups and increased sensitivity to the pain of the ingroup manifested as greater empathic concern for the ingroup who went through negative racial experiences (Dovidio et al., 2010).

However, other studies (Bruneau, Dufour, & Saxe, 2012) that compared empathy for ingroups versus conflict and neutral outgroups, found that mPFC was active more strongly both when observing ingroup and conflict outgroup pain compared to neutral outgroup pain. Conflict outgroups are more relevant for defining the group identity in the first place and represent a threat to that identity; thus this finding suggests that mPFC activity reflected personal closeness and relevance, and not simple similarity and membership. In summary, studies on emotional pain also point to intergroup empathy bias in brain responses, but leave the question of the influences shaping the mechanism and intensity of the bias unresolved.

Intergroup empathy bias was shown to be sensitive to contextual variables, as demonstrated in intervention studies whose goal was to reduce the bias in brain responses. For example, intergroup empathy bias was smaller if we had previously primed an independent versus interdependent self-concept in participants (Wang et al., 2015, but see Jiang, Varnum, Hou, & Han, 2014). The intensity of intergroup empathy bias was also responsive to the request to individualise

other-race members (Sheng & Han, 2012; Sheng, Liu, Li, Fang, & Han, 2014), as well as the amount of contact participants had with racial outgroups (Cao et al., 2015); however, the bias increased if participants were reminded of their own mortality (Li et al., 2015).

The findings about the effectiveness of dividing participants into minimal groups for reducing bias are not unanimous. When the new, minimal group membership is orthogonal to the racial group membership, shared membership in minimal groups is expected to reduce the empathy bias via recategorizing the racial outgroup members as minimal ingroups. In some studies, this was the case (Shen, Hu, Fan, Wang, & Wang, 2018; Sheng & Han, 2012). However, some studies do not find a significant effect of minimal groups over racial (Cao et al., 2015; Contreras-Huerta, Baker, Reynolds, Batalha, & Cunnington, 2013; Contreras-Huerta, Hielscher, Sherwell, Rens, & Cunnington, 2014), although the implicit measures of preference did point to a positive evaluation of the new ingroups (both racial ingroups and racial outgroups), suggesting that the minimal group manipulation was successful. These studies indicate that race automatically shapes empathic neural responses.

In summary, studies investigating interventions to reduce intergroup empathy bias in neural responses are heterogeneous and also differ by the mechanism they aim to employ. For example, priming independent self-construal aims to question and decrease the importance of group identification and thus decrease the salience of the (out)group identity of the target. Classifying participants into minimal groups orthogonal to group membership has the exact opposite goal: these studies rely on incorporating outgroup members into a new group identity and expect them to be seen as ingroups. Both of these strategies were borrowed from social identity theory (Tajfel & Turner, 2004), and were expected to improve the evaluation of the outgroup, albeit through a different mechanism. However, if we aim to promote equality in an intergroup context, these mechanisms have dramatically different implications for intervention design in the real world. Hence it is important to analyse these interventions with respect to the process of creating bias they aimed to interfere with as well as the specific empathic response they targeted.

### *Biased Empathic Responses to Ingroups Versus Non-racial Outgroups*

Behavioural studies on empathic responses to non-racial ingroups and outgroups are not very numerous. In addition, they are heterogeneous concerning the groups studied: sexual groups, cultural groups, university affiliation, and minimal groups.

Stürmer and collaborators (Stürmer, Snyder, Kropp, & Siem, 2006; Stürmer, Snyder, & Omoto, 2005) studied cultural groups and sexual groups. In these studies, baseline intergroup empathy bias was not registered: there were no differences in initial empathy towards ingroups compared to outgroups. However, empathy was shown to moderate intergroup helping. Specifically, empathy predicted readiness to help when the target's sexual orientation (Stürmer et al., 2005) and cultural group membership (Stürmer et al., 2006) matched the participants'.

In contrast, studies that used university affiliation as criteria for group differentiation (Tarrant, Dazeley, & Cottom, 2009), as well as minimal group studies (Montalan et al., 2012, Cikara, Bruneau, Van Bavel, & Saxe, 2014), pointed out that social categorization influenced the baseline empathy towards ingroups and outgroups. However, these studies do not agree on the form of bias. For example, in a series of experiments by Cikara and collaborators (Cikara, Bruneau, Van Bavel, & Saxe, 2014) outgroup hate was identified as the dominant form of bias, as well as that competitive circumstances increased it. Participants were divided into minimal groups (ostensibly based on a fake personality test, in fact, they were divided randomly). In this study, authors measured the participants' empathic responses to positive and negative events happening to ingroup

and outgroup members (e.g. S/he sat on a chewing gum on the bench). Participants displayed more empathy for the ingroup and more counter-empathic responses for the outgroup, especially in competitive circumstances, and including a third, neutral outgroup pointed to decreased reactivity to outgroups ("outgroup hate") as the mechanism. In contrast, intergroup empathy bias was registered in a behavioural study assessing physical pain (Montalan et al., 2012), but in its ingroup favouring form. In this study, the participants were classified into minimal groups and after that, they assessed the intensity of the pain in painful and painless images of hands and feet, while instructed to imagine those limbs belonged to themselves, to minimal ingroups, or minimal outgroups. It was shown that participants took longer to assess the painfulness of painful versus non-painful images when imagining themselves or the ingroup but not when imagining the outgroup member. This longer assessment however resulted in higher painfulness estimates only when comparing the painful-neutral difference between ingroup and outgroup but not the self condition. In addition, dispositional empathy predicted how long participants were going to spend on assessing ingroup pain. Contrasting previously described studies, this one pointed to "ingroup love" as the dominant mechanism, but also indicated that dispositional empathy and not our evaluations of ingroups and outgroups (i.e. identity-related measures) predicted the susceptibility to ingroup empathy bias.

Neuroscience studies of intergroup empathy bias are also heterogenous with respect to the groups studied – stereotype-defined groups (Cikara & Fiske, 2011), university affiliation (Richins, 2017), and fan identity (Hein, Silani, Preuschoff, Batson, & Singer, 2010; Cikara, Botvinick, & Fiske, 2011). What is common in all studies is that they all directly or indirectly studied competitive groups or competition-potential (i.e. status) ranked groups. Before describing them individually, we will comment on and problematise the Stereotype content model (SCM; Fiske, Cuddy, Glick, & Xu, 2002) as the main criteria for designating ingroups and outgroups. This line of research is interesting because the division into us and them is operationally defined by grouping the targets based on their stereotypical traits, i.e. individual internal characteristics. That is, internal traits are the reason and not the consequence of categorization; it can be argued that the criterion for the division is also similar to the previously discussed Stürmer et al., 2005 study. According to this model, the world is not divided only into us and them, but they represent a diverse and differentiated category. Beliefs about groups are organised along two dimensions: warmth and competence, which results in four categories of stereotypes about other groups, expected to shape different responses to different targets. Warm and competent groups elicit pride and cold and incompetent disgust, and groups from the remaining quadrants are ambivalent – we pity the warm and incompetent groups and envy the cold and competent ones. The main idea behind the Stereotype content model – that specific beliefs about certain groups evoke specific and distinguishable reactions – is in line with studies pointing out that the specific stereotypes about Black individuals are to blame for the pain treatment disparity, and not the outgroup status per se. The main issue with SCM studies is that they presuppose but do not assess that the participants identify with the warm and competent groups. In addition, it is questionable if these studies refer to social groups or social categories as abstract collections of individuals grouped by a characteristic they are perceived to possess by other people, but a characteristic that does not present a basis for self-identification. In terms of social identity theory, external but not internal criteria for the existence of a group are satisfied (Tajfel & Turner, 2004). Therefore, it is problematic to designate the differential responses as *ingroup* or *intergroup* bias in the absence of self-recognition of group membership.

Nevertheless, we will describe the SCM studies as they are frequently discussed along with ingroup empathy bias. In a series of neuroscience studies, Cikara and Fiske (2011) investigated if verbally presented positive, negative, and neutral events will cause different empathic reactions when happening to persons from different SCM quadrants (both behaviourally and neurally), and if these reactions can influence our readiness to harm those targets. The competence dimension can be discussed as a proxy for competitiveness (as competitive potential), although in SCM only the combination of high competence with low warmth (i.e. envy targets) is considered stereotypically

competitive (as competitive intention). It was demonstrated that in explicit responses participants differentiated between envy and pity targets – they had more compassion for the latter and were less ready to endorse harming them. However, willingness to harm both types of target was better predicted by neural compared to behavioural signals.

The idea that empathy bias depends on the specific outgroup defined by our beliefs about them was also explored in a thesis (Richins, 2017), in which the author used university affiliation to define group membership. In a series of experiments, it was recorded that self-reported empathy for pain depended on the number of criteria that made us different from them – the more differences, the less empathy. On the other hand, bias in neural responses was detectable only for competitive outgroups. The author pointed out that not all outgroups were considered equal, as well as that the differences can affect behavioural and neural responses differently.

Other neuroscience studies used fan affiliation to designate ingroups and outgroups. This line of division is explicitly based on competition between groups. For example, football fans were watching their ingroup or rival team outgroup receiving a painful shock, and they could choose if they wanted to help or not. In this study, both behavioural and neural measures pointed to intergroup empathy and schadenfreude bias, which predicted the willingness to help the ingroup and not help the outgroup (Hein, Silani, Preuschoff, Batson, & Singer, 2010). Similar results were obtained from baseball fans after they were verbally presented with the successes and failures of their own and rival team (Cikara, Botvinick, & Fiske, 2011). Subjectively negative outcomes (own team defeat or rival team victory) activated the pain matrix, which was related to the explicit assessment of the "painfulness" of those outcomes, and subjectively positive outcomes (own team victory or rival team defeat) activated the reward system, which was related to the self-reported probability of aggressive behaviour towards the rival.

**Summary**

This list of studies we reviewed did not aim to be exhaustive but to provide an insight into the diversity of the field of study. The results are in general agreement about one thing: intergroup empathy bias was identified in empathic reactions of different levels, from motor resonance with the physical pain of another to choosing to receive the shock instead of them. However, the studies differ in almost everything else. We will list several issues preventing us from reaching substantial conclusions about the intergroup empathy bias:

1. There is no agreement on the definition of empathy – neither theoretical nor operational. Consequently, there is heterogeneity in intergroup empathy bias research.

Behavioural studies of intergroup empathy bias are not numerous, they dominantly rely on self-report and are focused on higher-level empathic reactions such as empathic concern and compassion, as well as our readiness to react in a prosocial manner. The latter is arguably not empathy (whichever definition of empathy we endorse) but a hypothesised consequence of empathy, which illustrates how outstretched the concept is.

On the other hand, the majority of neuroscience studies elicit empathy with physically painful stimuli, thus investigating a lower-level empathic response. These studies are inconclusive about the relationship between biased neural responses they detected and explicit empathy bias (in behavioural measures), as well as with trait empathy and explicit and implicit prejudice. In addition, these studies are dominantly neuroscientific studies in focus, and socio-psychological constructs in studies of bias are discussed with respect to their neural foundations, and not vice versa.

Intergroup empathy bias is an empirical concept, i.e. a collection of empirical findings examining a range of empathy-related responses depending on the group identity of the target. As we have previously pointed out, the concept of intergroup empathy bias seems to primarily rely on neuroscientific studies, as reflected both in reviews and in the sheer number of neuroscience studies compared to behavioural. However, neuroscience studies can only investigate a limited range of phenomena listed under empathy; still, the general term empathy is used. The actual empathic responses measured in behavioural and neuroscience studies come from fundamentally different theoretical and methodological traditions and carry many implicit hypotheses about the phenomenon with their choice of measurement. In addition, each of those empathic responses is reasonably expected to engage a partially non-overlapping set of processes – it is very unlikely those will all be sensitive to social context to the same degree and manner.

2. Most studies were conducted on racial groups. Non-racial lines of differentiation in studies varied from minimal groups to sexual identities. The former always bring along several important characteristics that could shape bias and are impossible to separate from racial identity, which hinders our ability to make reliable inferences about the causes of bias. In addition, the social desirability explanation is difficult to exclude when investigating race. The studies on non-racial groups are few and very divergent with respect to theoretical rationale and criteria for division into us and them.

Therefore, we identify a need to systematically analyse and review intergroup bias studies to detect patterns and regularities from a set of divergent stories that will help us understand how, when, and in which responses intergroup bias emerges, what are the influences that shape it if it is possible and feasible to counter it and how to do it.

The most influential theoretical account of ingroup bias in social psychology, the social identity approach, views ingroup bias in "empathy" (as a general term) as a primary form of ingroup favouritism. However, bias is not defined as a blindly automatic phenomenon, but a response that emerges in the social context in relevant intergroup situations (Tajfel & Turner, 2004) and is functional with respect to situationally defined group goals. It implies both individual and contextual variability with respect to social variables influencing our individual propensities and shaping the demands of the situation. In our view, any discussion of social biases is incomplete without accounting for the context in which it emerges and the group-based goals it serves. This is in stark contrast with the view of empathy as an automatic response, implicated in neuroscience studies that view empathy as a neural response to empathy-eliciting stimuli; however, it is our impression that the concept of intergroup empathy bias is primarily based on neuroscience studies of physical pain.

The term social context is also critically undefined, though. In this thesis, we focused on the immediate context surrounding the empathic response within a specific study. As bias occurs and develops in the social context, we were interested in the empathic reactions embedded in specific empathy tasks. We were primarily concerned with how empathy was elicited and measured. Therefore, in the first part of this thesis, we reviewed and analysed the concept of intergroup empathy bias defined as differential empathic responses to ingroups and outgroups with a primary focus on the operational definitions of the study subject.

The second part of the thesis represents an empirical contribution to studying empathic reactions to physically painful stimuli. In keeping with the theoretical view that the social contextual variables are crucial for the emergence of ingroup bias, we contributed to our knowledge about the intergroup empathy bias for physical pain by conducting a series of behavioural studies that varied in ecological validity with respect to immediate contextual demands of the empathy-eliciting tasks. Specifically, we (a) replicated the most frequent empathy measurement paradigms from neuroscience studies we identified in the review to identify if intergroup empathy bias can be behaviourally detected when social desirability concerns were minimised and (b) assessed empathic reactions to pain elicited with progressively more ecologically valid and socially contextualised

tasks, to broaden the scope of empathic responses explored. These tasks varied in how well they imitated real responses participants could be prompted to give as group members outside of the experimental situation. These tasks also differed in their contextual relevance to the group identity in question, i.e. the degree to which the situation was truly social and the group identity was implicated in the assessment.

## Research Problem

The goal of this thesis was to contribute to our understanding of intergroup empathy bias in two ways.

(a) Via systematic review and analysis of studies that measured differential empathic response to ingroups and outgroups we aimed to explore the concept of *intergroup empathy bias,* i.e. to answer what was meant by *intergroup empathy bias* in current literature. More precisely, we wanted to analyse how empathy was theoretically and operationally defined in studies that focus on differential empathic responses to ingroup and outgroup members, as well as to summarise the results of those studies with respect to differences in study design.

(b) By conducting a series of studies on empathic responses to physical pain of ingroups and outgroups we aimed to address several limitations of neuroscientific studies of intergroup empathy bias for physical pain, in both theoretical and methodological aspects.

First, we are going to present and discuss the review study, and after that the experimental studies that were based on the review study to a variable degree. Specific goals for each study are presented in their corresponding sections.

## Intergroup Empathy Bias: An Overview

### Goals

The main goal of this study was to review and analyse intergroup empathy bias studies with respect to theoretical and operational definitions of empathy and bias.

Specifically, we aimed to:

1) Analyse neuroscience studies of intergroup empathy bias with respect to their results on behavioural indicators of empathy

   a) Analyse whether the conclusions derived from measures of empathic neural activity matched the conclusions suggested by the behavioural indicators

   b) Analyse how behavioural results were discussed in these studies, both in cases when their conclusions matched and when they didn't

2) For all studies that measured differential empathic responding to ingroups and outgroups, analyse the theoretical frameworks used to justify the research problem with respect to

   a) Socio-psychological theories and constructs explaining bias

   b) Theoretical definitions of empathy

3) For all studies that measured differential empathic responding to ingroups and outgroups, analyse how both empathy and bias were elicited and measured

   a) Count and summarise the tasks and measures used to elicit and operationally define empathy

   b) Count and summarise the categories used to evoke group identification, i.e. the division into ingroups and outgroups

The results of this systematic review were discussed in reference to three broad questions:

1) What is presently meant by "empathy" in intergroup empathy bias studies and how it was measured?

2) Which socio-psychological theories and concepts were called upon to justify the study of intergroup empathy bias or to explain it?

3) According to behavioural indicators, is there intergroup empathy bias?

### Procedure

We conducted the literature search during December 2021 in the databases ScienceDirect, GoogleScholar, and PsycInfo, using the following sets of search terms: intergroup/ingroup empathy bias/gap, parochial empathy, empathy + bias + race/ gender/ sex/ ethnicity/ nationality/ socioeconomic status/ sexuality/sports fans. As the search forms differ substantially among the databases, the exact search strategy for each of the databases is described in detail in Appendix A.

We identified records for retrieval by removing duplicates (i.e. records identified with more than one search string) and ineligible records and screening the remaining ones within each

database separately. After merging the resulting records and removing overlapping records (i.e. records identified in more than one database) and unavailable records, 262 records remained to be assessed for eligibility.

After excluding non-eligible records, intergroup empathy bias (IEB) studies on non-adult populations and studies where no information on the main effect of group identity on empathy measures is presented although it can be inferred from the study design and description that such information exists within the data, 63 journal articles remained with 88 studies.

Summary PRISMA2020 flow diagram (Page et al., 2021) is presented below, and PRISMA2020 diagrams for each database are presented in Appendix B.

Figure 1

*Summary Prisma2020 flow diagram for intergroup empathy bias studies*

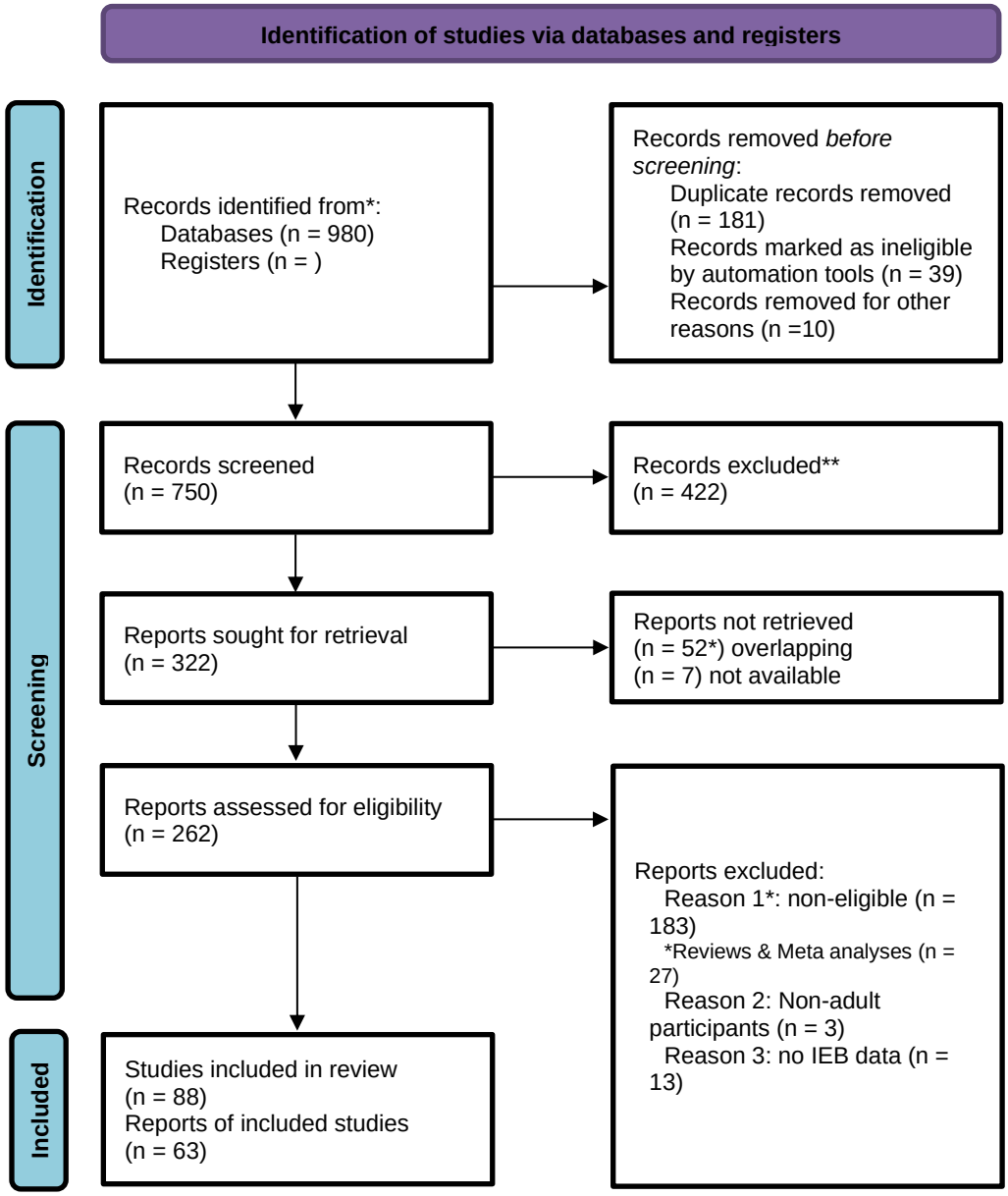


Table 1
*Articles and studies by study type: summary*

| Study type | N of articles | N of studies |
|---|---|---|
| behavioural | 26 | 44 |
| EEG | 5 | 5 |
| ERP | 13 | 15 |
| fMRI | 18 | 19 |
| tDCS&TMS | 2 | 2 |
| SCR | 2 | 3 |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response; articles could contain more than one type of study

Table 1 presents a summary description of articles and studies included in the analysis classified by study type. By *behavioural* we mean a study that assesses empathy only through behaviour willingly and consciously provided by the participant i.e. directly observable behaviour. We contrast these studies to *psychophysiological* studies that measure empathy indirectly i.e. via any psychophysiological response (recorded with specialised equipment from either the brain or body). [1] We chose to present and discuss them separately as we believe it is important to contrast how empathy is conceptualised within studies that rely on two fundamentally different methods of measurement.

We recorded several methodological and theoretical variables in each study:

1) General information about the study (authors, title, year)

2) General methodological information about the study

- Study type (psychophysiological (fMRI, EEG/ERP, TMS, tDCS, SCR) or behavioural)
- Study design (between subjects, within-subjects, mixed design)
- Sample (total number of participants, age, gender)

3) Information about empathy measurement

- Dependent variables (operational definition of empathy)
- Empathy-eliciting task
- Ingroup-outgroup division criteria

4) theoretical information about the study

- Socio-psychological theories and constructs referred to
- Explicit general definition of empathy authors referred to
- Definition of intergroup empathy bias

5) Behavioural results of the study

- Biased empathic response (yes/no)
- For neuroscience studies:

[1] Traditionally, in psychological literature, the phrase „behavioural measures“ explicitly excludes self-reports. However, in neuroscience, the phrase encompasses both self-reports and directly observable behaviours such as reaction time and these measures are contrasted to neural activity indicators. Our stance is that self-reports also represent a response to a stimulus in the person’s surroundings (a direct request to provide an answer) and therefore can be subsumed under the label „behavioural“.

- If there is a match between neural and behavioural indicators (yes/no)
- If and how the (mis)match was discussed

6) Additional measures in the study

- Dispositional empathy measures,
- Explicit indicators of general ingroup bias,
- Implicit indicators of general ingroup bias
- Other measures specific to individual studies

As physical pain is the main focus of this thesis and in line with the importance of physical pain as a basic human experience hypothesised to elicit empathic responses automatically and universally, we chose to present studies examining empathy for physical pain and non-physical empathy separately within both behavioural and neuroscience studies.

Within both physical pain empathy and non-physical empathy studies, categories of tasks, measures, and ingroup-outgroup division criteria were derived bottom-up, where it was possible to derive meaningful grouping criteria.

The articles in which empathy was assessed only behaviorally were typically multi-studies, whilst the articles in which empathy was assessed via psychophysiological activity were typically single studies. Considering the investment per participant (both temporal and monetary), and the fact that the vast majority of these studies were neuroscience studies, this statistic is hardly surprising but worth noting and will be discussed further. Here it is important to note that in extracting the theoretical information about the study, the unit of analysis was an article, that could but not always contain more than one study.

For theoretical positioning of empathy, we focused on the explicitly spelled out introductory definitions of the general notion, as well as explicit definitions of intergroup empathy bias ("Empathy is..."; "Intergroup empathy bias is...". Our initial idea was to analyse the theoretical derivation chain linking the general empathy definition to the specific empathic phenomenon in focus (in both theory and measurement), in order to understand how intergroup empathy bias fits within the broader theories on empathy. In other words, we expected to analyse how the authors viewed empathy bias in the context of their general understanding of empathy and the specific phenomenon of interest. However, this very soon proved to be completely unfeasible, as no explicit derivation chain could be detected in any of the articles.

In analysing the socio-psychological theoretical framework, we focused on the main rationale that justified the study design and/or used to explain the study results. In other words, we extracted the main theories, hypotheses, concepts, and empirical phenomena (i.e. regularities found in empirical studies) used to theoretically justify measuring empathy in the context of intergroup relations. In some cases, the authors explicitly named a theory they rely on to place empathy in the context of intergroup relations (e.g. *"we take an intergroup emotions theory perspective on empathy"*) or used theory-typical concepts along with citing seminal articles (e.g. *social identification*, with referencing e.g. Tajfel & Turner, 1979). In other cases, concepts were used without referencing a specific theory (e.g. *"empathy is a social emotion"*) or could be recognizably linked to broader approaches to studying social phenomena (e.g. motivated social cognition, dual-process theories). [2]

[2] For example, the authors of *Are liberals and conservatives equally motivated to feel empathy toward others?* (Hasson, Tamir, Brahms, & Cohrs, 2018) view empathy as a ***social emotion*** (they explicitly state it), subject to motivated regulation (contrasted to an automated and reflexive view on empathy). Some previous studies indicate that liberals are more empathic than conservatives; however, critics state most differences between liberals and conservatives are overinflated due to use of ideologically biased targets. Authors of this study view political ideology as both socially shared system of beliefs that reflects higher-order social goals which could shape our motivations to feel empathy; by

We now turn to a separate presentation of behavioural and psychophysiological studies. As our main goal was to understand what was implied versus what was actually measured in intergroup empathy bias studies, the disproportionate focus of the presentation will be placed on empathy measurement.

The complete database of articles discussed in this review is available at Open Science Framework (OSF; https://osf.io/gd5c8/).

## Behavioural Studies

We identified 44 behavioural studies in 26 articles, which were published between 1996 and 2021. A breakdown of study results by type of stimuli and line of ingroup/outgroup division is presented in Table 2.

Studies labelled *physical pain* (13 studies, 10 articles) focused exclusively on empathy for specific painful bodily events or states, while *non-physical pain* (31 studies, 16 articles) studies investigated a broader scope of empathic reactions.

Note that despite the *non-physical* label, the stimuli in some (but not all) of the studies contained descriptions of physical injuries or events that resulted in physical injuries. However, the main focus of those studies was not empathy for physical pain per se but more general empathic reactions towards bad events happening to others. We base this claim on the fact that in studies where both physical and non-physical bad events were presented separately, the empathic reactions were averaged across stimuli, and in studies describing events that resulted in physical injuries, those were only one of many aspects of the complex bad event happening to ingroups and outgroups.

Under non-physical empathy studies, we presented and discussed the studies we labelled *misfortunes studies* (17 studies, 6 articles) separately, as they use a common method of eliciting and/or measuring empathy, designed by M. Cikara (Cikara et al., 2014). These studies were either co-authored by M. Cikara and collaborators (13) or directly and explicitly (by authors' own writing) inspired by their work (4). Empathy was typically elicited by presenting a single-sentence description of a specific event or a misfortune (as the authors label it) happening to ingroups and outgroups and asking the participant to rate their congruent reaction to the event on a visual slider (variations and details will be described in detail within the specific section).

Thus, we will now present non-physical empathy studies first, followed by physical pain empathy studies. Within the former, we will separately discuss misfortunes studies (17) and other (14) studies.

---

that view liberals should be motivated to feel and actually feel more empathy across contexts (***motivated social cognition***). However, they believe that political ideology is also a common basis for social identification (referencing Tajfel & Turner, 1979) and as empathy improves social connections it could be expected that both groups are motivated to feel more empathy and actually feel more empathy towards their ingroups due to basic motivations to maintain and enhance positive group evaluation (***social identity theory***). Therefore, in a study that used ideologically equally relevant targets for both groups they test these hypoteses on a diverse sample of liberals and conservatives.

Table 2

*Ingroup empathy bias in behavioural studies by types of stimuli and types of groups*

| | stimuli | racial | | national | | political | | gender | | sexual | | religious | | minimal | | sport | | total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| non-physical empathy | questionnaire item | | | | | | | | | | | $1^{1}$ | | | | | | 1 | 0 |
| | vignettes | | | | 1 | 1 | | $1^{?}$ | 1 | | $1^{1}$ | $1^{1?}$ | | | | | | 3 | 3 |
| | ostracism | | 4* | | | | | | | | | | | | | | | 0 | 4 |
| | video | 1 | | | | | | | | | | | | | | | | 1 | 0 |
| | image | 1 | | | | | | $1^{?}$ | | | | | | | | | | 2 | 0 |
| | (mis)fortunes | | | 3 | | 2 | | | | | | | | 12 | | | | 17 | 0 |
| physical pain empathy | shoulder pain (med) | 2 | 2 | | | | | | | | | | | | | | | 2 | 2 |
| | real pain (med) | | | | | | | 1 | | | | | | | | | | 1 | 0 |
| | pain on hands and feet | | | 2 | 1 | | | | | | | | | 1 | | | 1* | 3 | 2 |
| | descriptions of pain | 1 | 1* | | | | | | | | | | | | | | | 1 | 1 |
| | facial expressions of pain | | 1 | | | | | | | | | | | | | | | 0 | 1 |
| | total | 5 | 8 | 5 | 2 | 3 | 0 | 3 | 1 | 0 | 1 | 2 | 0 | 13 | 0 | 0 | 1 | 31 | 13 |

Note. 1 – intragroup bias studies (i.e. within religious or sexual minority identity); ? – bias found in one but not all groups studied; * – outgroup bias; Y – bias registered; N – bias not registered

### *Non-physical Empathy*

**(Mis)fortunes.** ~40% of the behavioural studies assessed participants' empathic reactions to fortunes and misfortunes happening to ingroups and outgroups (17 studies).

A typical trial in these studies consisted of a sentence describing a specific event (*"Beth / Saana found out that her son was taking money without asking"/ "Ted / Adbullah found a sentimental possession he thought he had lost"*) and verbal (and sometimes also visual, e.g. a flag) cues of the target's identity (e.g. *"Beth is from USA/Saana is from Iraq"*). Participants were then asked to indicate how bad the negative event made them feel and how good the positive event made them feel on an anchored but unmarked slider. In two-thirds of these studies participants' counter-emphatic reactions (i.e. Schadenfreude – positive affect following negative events and Gluckschmertz – negative affect following positive events) were simultaneously assessed in the same manner (as separate ratings of incongruent reactions, i.e. "how good" for negative and "how bad" for positive events).

This description represents a prototype of the experimental procedure developed by M. Cikara and collaborators (Cikara, Bruneau, Van Bavel, & Saxe, 2014) which was used in all studies with some variations. For example, most of the studies presented both positive and negative events (10) while a minority presented only negative events (4) as empathy-eliciting stimuli. However, in 3 studies a positive or negative outcome (i.e. ostensibly winning or losing one trial in a competitive reaction time game) was considered a fortune/misfortune happening to self and teammates vs. members of the opposing team. In addition, empathy-eliciting stimuli were varied by their strength in one study (mild, e.g. *"Melanie stepped in dog poo"* vs extreme, e.g *"Melanie had a stranger throw up on her.")* or by adding short narrative descriptions about the targets before the events in two studies (e.g. *"Melanie was looking forward to going to college. Melanie lived in the city and was taking a bus to meet with an interviewer for one of the schools she applied to. She was anxious on the bus and didn't notice where she was going when she stepped off. Just off the bus, after she had slung her bag over her shoulder, Melanie stepped in some dog poo.")*. As for the empathy measurement, although the vast majority operationally defines empathy as a congruent emotional reaction, i.e. feeling good for positive events and feeling bad for negative events (15), in two studies, the participants were asked to indicate their general empathy towards ingroup/outgroup's misfortune on a Likert scale (1 – *none at all* to 5 – *a lot*).

All 17 studies were conducted on Mturk or another local survey company (12 on Mturk, 3 on Prolific, 1 on Kerdoivem, 1 on The Hellenic Research House) and usually represent a series of studies (6 articles total vs 17 studies). All studies used within-subject design. The number of participants ranged from 37 to 811, with $M \approx 215$ and $Mdn \approx 116$ participants.

12 out of 17 studies used the minimal group paradigm, while the remaining examined IEB on real groups – political (2) or national (3). Minimal groups were created by ostensibly classifying participants based on their personalities in 9 studies (participants rated their personality on NEO items) and in 3 studies they were randomly paired in a team with either a human or a robot. However, although these groups were not always put in direct competition for awards (i.e. the functional relations between groups were not always competitive), in all studies the groups were put in a competitive context (e.g. in one study the instructions explicitly state that a group's award or performance is independent of another group's, but their existence is created within a "problem-solving challenge" context which implies their performances could be compared).

All 17 of the 17 misfortunes studies find ingroup empathy bias i.e. more baseline empathy for the ingroup compared to the outgroup, that can be increased by stressing the competitive nature of the interactions and reduced by presenting individuating information about the targets or manipulating group entitativity. On the other hand, episodic simulation interventions, as well as informing the group of their better performance or the positive or negative outcome of the competition had no effects on the bias registered at baseline.

As previously mentioned, empathy definitions and theoretical reasoning in each article were common to all studies in that article; therefore the unit of analysis for empathy definitions and socio-psychological frameworks used as a theoretical rationale for the study design was an article.

Empathy was explicitly defined in all but one article (i.e. in 5 out of 6 total articles), as empathic concern (1) or congruent feeling (4), although its cognitive "counterpart" is sometimes mentioned in the introduction. As (in)congruent emotional responses were measured in most studies, we could state that the operational definition of empathy mainly corresponded to the theoretical (and sometimes directly replaced it).

Intergroup empathy bias was sometimes referred to as a "group-based emotion", and therefore implicitly embedded within a broader theory on biased emotional processing. However, most studies referred to intergroup empathy bias as an empirical fact. By "treated as an empirical fact" we mean that intergroup empathy bias was introduced as an empirical phenomenon reliably found in many studies. To illustrate, Cikara et al. (2014, p. 110) stated that: *"Despite its early*

*origins and adaptive functions, empathy is not a universal response. People often feel less empathy for strangers who belong to a different racial, political, or social group, compared to strangers who are described as belonging to the same group (Batson & Ahmad, 2009; Davis, 1994; Hornstein, 1978): we term this difference the intergroup empathy bias (Cikara, Bruneau, & Saxe, 2012)."* We argue that the existence of intergroup empathy bias was primarily claimed on the basis of neuroscientific studies of empathy, as the first three references offer conceptual arguments and insights into why it might be difficult to empathise with the outgroup but no empirical evidence and the last one summarises evidence of „dampened and disrupted empathy for outgroups" primarily in neuroscience studies. Cikara et al. (2014, p. 111) further state: *"As dozens of recent papers from social and developmental psychology and cognitive neuroscience demonstrate, dampened or absent empathic responses (and associated physiological indicators) are particularly likely for social or cultural out-groups (e.g., Avenanti, Sirigu, & Aglioti, 2010; Chiao & Mathur, 2010; Cuddy, Rock, & Norton, 2007; Decety, Echols, & Correll, 2009; Hein et al., 2010; Masten, Gillen-O'Neel, & Brown, 2010; Tarrant, Dazeley, & Cottom, 2009; Xu, Zuo, Wang, & Han, 2009)."* Among these, 5 out of 8 were neuroscience studies; among the remaining three, one did not measure empathy but helping behaviour (Cuddy, Rock, & Norton, 2007). The explicitly stated goal of the Cikara et al. (2014, p. 113) study was to further dissect intergroup empathy bias as *"a well-established phenomenon among real social groups and categories. Our goal is not merely to show that intergroup empathy bias manifests among novel groups as well, but rather to elucidate precisely the causal factors in intergroup contexts that drive empathy and counter-empathic responses."*

Social identity theory was most frequently used as theoretical rationale: it was the main theoretical framework in 4 out of 6 studies, and the first article using the misfortunes paradigm which included five separate studies was explicitly designed to systematically test various predictions stemming from this framework (effects of competition, ingroup love vs outgroup hate, group entitativity). In addition, concepts of humanization (2) and salience of group identity were additionally invoked to either justify the design or explain the results. Out of the remaining two articles, one was based on the intergroup emotions theory, and one article was theoretically focused on a non-social intervention designed to reduce the existing IEB, i.e. was not grounded in any specific socio-psychological theory.

Table 3

*Behavioural studies presenting (mis)fortunes*

| Title | Year | S/I | N | Stimuli | Group | Dependent variable | Bias (Y/N) | Empathy definition | IEB definition | Theories and concepts explaining IEB |
|---|---|---|---|---|---|---|---|---|---|---|
| Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses | 2014 | I | 49 | positive and negative events | minimal (falling behind) | congruent and incongruent emotions rating | y | congruent emotional response | IEB as an empirical fact | SIT; competitive threat |
| | | I | 202 | | minimal (competitive, cooperative, independent) | | y | | | SIT; competition |
| | | S | 48 | | minimal and distant/unaffiliated | | y | | | SIT; ingroup love, outgroup hate |
| | | I | 45 | | minimal (falling behind) | | y | | | SIT; competitive threat |
| | | I | 72 | | minimal (similar/dissimilar) | | y | | | SIT; group entitativity |
| Minding the Gap: Narrative Descriptions about Mental States Attenuate Parochial Empathy | 2015 | I | 322 | positive and negative events | minimal | congruent emotions rating | y | operationally defined as self-reported congruent emotion | operationally defined as difference in rating for IG and OG | SIT, salience, (de)humanization |
| | | I | 234 | | | | y | | | |
| | | I | 187 | | | | y | | | |

*Note*. S/I – simple/intervention study concerning IEB; N – sample size; y – yes; n – no; IG – ingroup; OG - outgroup

Table 3 (continued)

| Title | Year | S/I | N | Stimuli | Group | Dependent variable | Bias (Y/N) | Empathy definition | IEB definition | Theories and concepts explaining IEB |
|---|---|---|---|---|---|---|---|---|---|---|
| Parochial Empathy Predicts Reduced Altruism and the Endorsement of Passive Harm | 2017 | S | 502 | negative events | national/conflict (Hungarian – Refugees) | general empathy rating | y | empathic concern | empathy as group-based emotion; IEB as an empirical fact | intergroup emotions theory, ingroup love vs. outgroup hate SIT |
| | | S | 467 | | national/conflict (Greeks - Germans) | | y | | | |
| | | S | 84 | positive and negative events | national/conflict (USA - Arabs) | congruent and incongruent emotions rating | y | | | |
| Activating episodic simulation increases affective empathy | 2021 | I | 811 | negative events | political (Democrat/ Republican) | congruent emotions rating | y | affective empathy (congruent feeling) | IEB as an empirical fact | / |
| | | I | 199 | | | | y | | | |
| The Opposite of Love: How parasocial interaction with NPCs can be enhanced through intergroup competition | 2020 | | 116 | positive and negative events | minimal | congruent and incongruent emotions rating | y | / | IEB as an empirical fact | SIT; parasocial interaction |
| Empathy and Schadenfreude in Human–Robot Teams | 2021 | S | 37 | positive and negative outcomes | minimal; species? | congruent and incongruent emotions rating | y | cognitive and affective; vicarious feelings | IEB as an empirical fact | SIT; competition; humaneness |
| | | I | 87 | | | | y | | | |
| | | I | 93 | | | | y | | | |

*Note*. S/I – simple/intervention study concerning IEB; N – sample size; y – yes; n – no; IG – ingroup; OG - outgroup

**Miscellaneous Studies**. In about one-third (14) of the remaining non-physical pain studies presented in 10 separate articles, participants were exposed to empathy-eliciting stimuli that were very variable in content: verbal scenarios (6) of negative events resulting in general negative outcomes i.e. physical and/or emotional injuries (people being raped, beaten, falling ill, going to prison), videos of persons describing psychological distress or images contextually implying complex emotional states (3) or observing events where people were excluded from social interactions (ostracism) (4). In one study participants were told to imagine an ingroup or an outgroup while answering the questions, i.e. there was no specific empathy-eliciting task. In all studies, participants were asked to rate the extent they feel empathy for the target on a Likert scale. The ratings were either single-item general empathy ratings (5 studies) (*to what extent do you empathise?/are you able to empathise*), single-item empathy components ratings (1) or represented average scores of 3-4 empathy-related ratings (6), with two exceptions which used a standardised empathy test (MET-core-2, Drimalla et al., 2019; & State Empathy Scale, Shen, 2010). Empathy targets varied by race (6), gender (3), sexual (1), political (1), national (1), or religious (2) orientation. Most studies were simple studies with respect to empathy i.e. did not aim to influence or manipulate the level of empathy towards ingroups and outgroups. Two studies did try to influence empathic responses towards the targets: one intervention study manipulated the salience of rape myths within vignettes while the other study compared the participant responses when they were instructed to put themselves in the target's shoes (empathy induction intervention) vs no instruction.

The number of participants per study ranged from 56 to 1958, with M ≈ 619 and Mdn ≈ 277 participants. However, it should be noted that 6 studies conducted on Mturk had M = 969 participants, while other studies conducted on helping professionals (1), students (4), or specific subpopulations had significantly fewer participants.

Before continuing, we would like to remind the reader that the unit of analysis in further text is an article, which could contain more than one study.

In one-third of the articles (3), empathy was not explicitly defined. The remaining two-thirds defined it somewhat variably, but the emotional quality of empathy was present in each definition (7) and the cognitive counterpart (5) in most of them. The measured construct sometimes roughly corresponded to the definitions (e.g. an article defining empathy as empathic accuracy and affect sharing measured both accuracy and sharing), but this correspondence was very loose, general, and simplified in most studies (e.g. an article defines empathy as consisting of affect sharing and understanding that leads to sympathy and compassion, but operationally defined the empathic response as an average rating of empathy, sympathy, and compassion).

Intergroup empathy bias was explicitly mentioned in 8 out of 10 articles. In all of them, the phenomenon is treated as an empirical fact, i.e. presented as a list of studies in which its existence had been previously established. Here is an illustration of a typical way to introduce and summarise the concept of intergroup empathy bias: *"Most notably, ample research suggests an "empathy gap" across group boundaries whereby individuals are less likely to empathize with outgroup than ingroup members in a range of different circumstances, such as when observing another person's expressions of sadness (Gusell & Inzlicht, 2012), witnessing another individual's experience of physical pain (Avenanti, Sirigu & Aglioti, 2010, Xu et al., 2009), or learning of negative events or misfortunes another person has experienced (Bruneau, Cikara & Saxe). Deficits in empathy have been identified across various naturally occurring intergroup boundaries (e.g., political or ethnic differences) as well as boundaries created in the laboratory, indicating that an empathy gap for outgroup members is a reliable and robust phenomenon."*, Petsnik & Vorauer, 2020, p. 3). We want to point out that studies in this study group either refer to misfortunes studies or neuroscience studies to explain the phenomenon of intergroup empathy bias.

It is important to note that only four out of ten articles directly investigated empathy and intergroup empathy bias (all of these define empathy explicitly). Two of these four articles were theoretically embedded within the motivated empathy account (Zaki, 2014). In the first one, the study authors claimed that "the broader phenomenon of intergroup empathy bias" (i.e. IEB as an empirical fact) indicated that empathy was susceptible to contextual influences and that people were likely to be motivated to empathise with ingroups whom they favour – this was contrasted to the notion that liberals were more empathic in general than conservatives because empathy is in line with their social goals. The topic of the other study was trait empathy[3] and IEB as a collection of studies was used to illustrate its expected parochial consequences, i.e. to justify the expectation that trait EC will amplify value-led or motivated responses to rape vignettes. The third study discusses empathy bias within a broader context of ingroup biases stemming from social categorization. The fourth study contrasted the hypotheses that intergroup empathy bias could be explained by better recognition vs physical similarity with the ingroups and interpreted its results within the Perception-Action Model (Preston & De Waal, 2002).

In another four articles authors embedded empathy measures towards ingroups and outgroups within a specific research question hypothesised to be related to empathy and intergroup empathy bias or mediated by it. In all of these studies intergroup empathy bias is treated as an empirical fact, i.e. as a research phenomenon consistently appearing as measured by several methods. For example, in one of these articles, authors studied how dominant group members react to dominant-on-dominant vs dominant-on-disadvantaged ostracism and hypothesised that based on IEB as a phenomenon we could expect that reactivity to ostracism of disadvantaged outgroups compared to ingroups will be reduced. Similarly, intergroup empathy bias was used to justify the hypotheses about altruistic responses in traumatised social groups, victim group responses to hate crime, and intragroup religious conflict over alterations of religious rituals due to the COVID-19 pandemic.

The two remaining articles do not refer to IEB as a concept at all. One of these studies is a general study of bias that included an empathy rating among others and was the only one conducted during the 1990s. The other study investigated the influence of sex hormones on emotional processing in women and was included in the review because it fulfilled the criteria we listed (it measured empathy and the targets varied by gender) but came from a different research domain.

As for socio-psychological theoretical rationale, social identity approach (SIT & SCT) seminal studies were directly referenced in 3 articles and the concepts of social identity and social categorization as well as related specific concepts (social identity threat, black sheep effect, etc) were used as theoretical rationale or post hoc explanations in 3 additional articles. However, we want to stress that 2 of these articles referencing SIT&SCT directly investigated empathy bias and contrasted the consequences of motivated social cognition because of different values perceivers hold to their group membership per se (both of these saw empathy as a motivated phenomenon). One article was dominantly based on Intergroup emotions theory (although referencing SIT and recognizing SIT and IET as related theories, as well as exploring several specific concepts in addition). Perceived similarity without referencing a specific theory was drawn as a main explanation in one article. Another article invoked physical similarity to explain the results, but no concept from social psychology was used as a theoretical rationale, and one article was thematically unrelated to social psychology.

Out of 14 studies in 10 articles, in 7 studies the authors observed ingroup empathy bias, out of which 3 studies observed bias only in one but not the other group in the sample (e.g. ingroup bias found only in males but not females). Seven studies find no empathy bias.

[3] The authors measured state empathy (labelled responses) as well, hence this article entered our review.

Table 4

*Behavioural studies presenting miscellaneous content*

| Title | Year | S/I | N | Design | Stimuli | Group | Dependent variable | Bias (Y/N) | Empathy definition | IEB definition | Theories and concepts explaining IEB |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Psychotherapist Bias With Hispanics: An Analog Study | 1996 | S | 56 | bs | video | racial | single rating (not at all-easily able to empathise) | y | / | / | empirical evidence of bias; similarity; anticipation of negative affect |
| Empathy towards individuals of the same and different ethnicity when depicted in negative and positive contexts | 2013 | S | 198 | ws | picture | racial | separate ratigns (difficult/easy to take perspective/ feel the target's feeligns/ understand) | y | ability to share, perceive or imagine experiences of others; PT, emotional simulation, emotional regulation | IEB = a collection of empirical findings; PAM | recognition ability vs similarity; interpreted within the PAM |
| On the Ironic Effects of Being Empathic: Consequences for Attitude Polarization and Intergroup Conflict (UNPUBLISHED THESIS) | 2017 | I | 253 | ws | verbal scenario | gender | average rating (empathy, sympathy, compassion, support) | n | empathic concern (state vs trait) | IEB = a collection of empirical findings; motivated empathy | motivated social cognition ("vested interest" - values vs group membership per se), role of trait EC in polarization |
| | | S | 539 | ws | | | | y | | | |
| Are Liberals and Conservatives Equally Motivated to Feel Empathy Toward Others? | 2018 | S | 910 | bs | verbal scenario | political | average rating (empathy, sympathy, compassion – motivation and actual) | y | affect sharing, understanding -> sympathy, compassion | IEB = a collection of empirical findings; motivated empathy | motivated social cognition (political ideology, group membership per se - SIT, empathy as social emotion |
| Understanding victim group responses to hate crime: shared identities, perceived similarity and intergroup emotions | 2018 | S | 197 | bs | verbal scenario | sexual | average rating (sadness, sympathy, respect, empathise) | n | congruent, other oriented emotional response | IEB = a collection of empirical findings | integroup emotions theory (Mackie, Devos & Smith, 2000), social identity theory (Tajfel & Turner, 1986); IEB as mediator between group-based identity and group-based reactions; perceived similarity; common ingroup identity; victim blaming |
| Utjecaj etničke pripadnosti aktera u scenariju na samoprocjenjenu empatiju (UNPUBLISHED THESIS) | 2020 | I | 195 | bs | verbal scenario | national? | average rating (State empathy scale, Shen, 2010) | n | perspective taking (associative), emotion sharing (affective), understanding (cognitive) | IEB = a collection of empirical findings | ingroup biases as consequence of social categorization; absence of bias in results explained post hoc by individuation, black sheep effect, lack of categorization as ingroup/outgroup members |

*Note*. S/I – simple/intervention study concerning IEB; N – sample size; y – yes; n – no; IG – ingroup; OG – outgroup; PAM – Perception-action model

Table 4 (continued)

| Title | Year | S/I | N | Design | Stimuli | Group | Dependent variable | Bias (Y/N) | Empathy definition | IEB definition | Theories and concepts explaining IEB |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Do dominant group members have different emotional responses to observing dominant-on-dominant versus dominant-on disadvantaged ostracism?... | 2020 | S | 924 | bs | ball tossing game (ostracism) | racial | single rating (to what extent do you empathise) | n | / | IEB = a collection of empirical findings | reduced reactivity hypothesis: motivation to maintain positive social identity; IEB and bystander apathy as phenomena; enhanced reactivity hypothesis: social identity threat via violation of contemporary norms, social dissimilarity rule in ostracism reactions as phenomenon |
| | | S | 1958 | bs | | | single rating (to what extent do you empathise) | n | | | |
| | | S | 1230 | bs | | | average rating (moved, compassionate, warm, soft-hearted, sympathetic, and tender) | n | | | |
| | | S | 301 | bs | | | single rating (to what extent do you empathise) | n | | | |
| Alterations in Religious Rituals Due to COVID-19 Could Be Related to Intragroup Negativity: A Case of Changes in Receiving Holy Communion in the Roman Catholic Community in Poland | 2021 | S | 179 | ws | verbal scenario | religious | EC (compassionate, softhearted, moved, and warm) and distress (Upset, distressed, worried, and troubled) | y | / | IEB = a collection of empirical findings | social identity theory, IEB as a special case of general biases, intragroup conflict |
| Empathy-Mediated Altruism in Intergroup Contexts: The Roles of Posttraumatic Stress and Posttraumatic Growth | 2021 | S | 1660 | bs | questionnaire item | religious | single rating (not at all – extremely feel empathy) | y | other-directed emotional response (sympathy, tenderness, compassion); cognitive vs emotional empathy | IEB = a collection of empirical findings | empathy – prosocial behaviour relationship, moderating role of posttraumatic stess/growth, IEB as cause of less motivation to alleviate outgroup's distress |
| Sex-hormone status and emotional processing in healthy women | 2021 | S | 72 | ws | picture | gender | standardised test (cognitive, affective) | y | identify and co-experience other's emotional states and respond appropriately | / | / |

*Note*. S/I – simple/intervention study concerning IEB; N – sample size; Y – yes; N – no; IG – ingroup; OG – outgroup; PAM – Perception-action model

### *Physical Pain Empathy*

Contrary to our preliminary review of ingroup empathy bias literature, we found 13 behavioural studies examining empathy for physical pain, i.e. presenting exclusively physically painful stimuli. Five out of thirteen studies examined pain in the medical context. These studies used videos (3 studies, assessing empathic concern) or pictures (1, asking the participants to rate the pain intensity) of persons with painful expressions while performing shoulder movement as empathy-eliciting stimuli (and they were ostensibly referred to a medical professional due to shoulder pain). However, one study used real pain in medical procedures to investigate bias (bias was defined as a difference in pain estimates between the patient and the physician). Another five studies out of thirteen presented participants with images of hands and feet ostensibly belonging to ingroups and outgroups in painful and neutral situations and they operationally defined empathy as pain intensity ratings (either Likert or VAS). Group identity of the target was designated by presenting the participants with the typical names of the ingroups or the outgroups (3) or their characteristic visual symbols (1) before the painful events, or they were asked to imagine (1) themselves, the ingroup and the outgroup experiencing the visually presented painful event. The remaining studies asked the participants to imagine verbally described painful events happening to both themselves and the person in the photograph and rate the intensity of the pain (2), or rate pain intensity and self-unpleasantness of painful facial expressions of ingroups and outgroups (1).

Pain studies were conducted on helping professionals (3), students (4), the general population or an identity-defined subpopulation (4), or a combination of Mturk and studentpopulations (1) (one study did not specify the sample type). Racial groups were most frequently investigated (7), but several studies investigated national/conflict groups (3); in the remaining studies, the line of division was gender (1), fan identity (1), or minimal (1). The number of participants ranged from 20 to 240, with $M \approx 76$ and $Mdn \approx 55$ participants.

In total, in seven studies the authors reported ingroup empathy bias, in four they found null effects and in two they found outgroup empathy bias.

Empathy was explicitly defined in six out of ten articles and all definitions included both its affective and cognitive qualities. As mentioned, empathy was most frequently operationally defined through pain intensity ratings (9 out of 10 articles). Pain intensity assessment could theoretically rely on both vicarious emotional reactions and cognitive analysis, therefore in a broad sense we could say that the operational definition overlaps with the theoretical. However, we argue that taking pain intensity ratings as indicators of empathy implicitly relies on the hypothesis that they are influenced by the amount of vicarious sharing, which will be problematised in further text.

As in non-physical pain studies, intergroup empathy bias was presented (in nine out of ten articles where it was explicitly mentioned) as an empirical phenomenon, i.e. through a list of mainly neuroscience and misfortunes studies that had found less empathy for the outgroup. In two of these articles, empathy bias was investigated in the context of pain treatment bias, i.e. the authors were primarily interested in its consequences. In the remaining seven studies intergroup empathy bias per se was the main topic: if it could be elicited in minimal groups (1), was it sensitive to various contextual information (3) or physiological states (2), or how it could be predicted (1) and how it could be explained (1)[4].

Social identity approach and theory-related concepts were mentioned in the theoretical rationale of four articles (out of ten); however, only two of these four were based primarily on these theories and they were explicitly referenced. In the other two articles, the concepts of social identification and categorization were used to explain why the authors expected specific

[4] One study deals both with predicting and modifying ingroup empathy bias, hence the sum of the numbers in brackets is eight and not seven.

interventions they employed were expected to reduce intergroup empathy bias. [5] Various effects of oxytocin were the main topic of two out of ten articles and its hypothesised effects on empathy were expected through manipulating social salience. Importantly, in these two articles (and one more which was based on the main idea of the social identity approach) members of groups in acute intractable conflict participated in the study and this enabled the authors to investigate whether the main mechanism of bias was ingroup love or outgroup hate by contrasting participants' reaction to conflict vs neutral outgroups. Three additional articles relied on specific concepts to explain the results: two articles primarily relied on specific stereotype content and one more stressed the importance of gender norms and similarity of experiences. The last of the remaining articles listed evidence of racial bias in medical treatment but the authors did not employ any specific concept or theory to explain the existing bias.

[5] In one of these studies authors investigated the relationship of IEB with dehumanization and SDO and whether IEB could be reduced with signaling the target's multiple identities, and the other study primarily focused on reducing IEB through residential mobility manipulation (its expected effectiveness was justified with multiple theoretical concepts).

Table 5

*Behavioural studies presenting physically painful content*

| Title | Year | S/I | N | Design | Stimuli | Group | Dependent variable | Bias (Y/N) | Empathy definition | IEB definition | Theories and concepts explaining IEB |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reducing racial disparities in pain treatment: The role of empathy and perspective-taking | 2011 | s | 40 | ws | painful shoulder movement | racial | empathic concern scale | n | cognitive and emotional process that fosters emotional understanding; linked to prosociality | IEB as an empirical fact | / |
| | | s | 51 | ws | | | | y | | | |
| | | I | 60 | ws | | | | y | | | |
| Behavioural investigation of the influence of social categorization on empathy for pain: a minimal group paradigm study | 2012 | s | 36 | ws | painful events on hands and feet | minimal | pain intensity rating | y | affective (sharing), cognitive (understanding) | IEB as an empirical fact | SIT (mere categorization effect) |
| Racial Bias in Perceptions of Others' Pain | 2012 | s | 35 | bs | verbal descriptions of painful events | racial | pain intensity rating | n | | IEB as an empirical fact | stereotypes and specific stereotype content |
| | | s | 240 | bs | | | | y | | | |
| Giving peace a chance: Oxytocin increases empathy to pain in the context of the Israeli—Palestinian conflict | 2013 | i | 55 | ws | painful events on hands and feet | national/ conflict, neutral | pain intensity rating | y | / | IEB as an empirical fact | salience, intergroup conflict, ingroup favouritism vs outgroup derogation |
| The role of oxytocin in empathy to the pain of conflictual out-group members among patients with schizophrenia | 2014 | i | 27 | ws | painful events on hands and feet | national/ conflict, neutral | pain intensity rating | n | affective response requiring understanding; EFP – automatic distress when witnessing pain | IEB as an empirical fact | salience, intergroup conflict |
| Predicting and Intervening Intergroup Empathy Bias in Soccer Fans (UNPUBLISHED THESIS) | 2016 | i | 89 | ws | painful events on hands and feet | identity | pain intensity rating | n | affective, cognitive | IEB as an empirical fact | dehumanization, SDO; multiple categorization intervention: social categorization theory, |
| The impact of implicitly and explicitly primed ingroup-outgroup categorization on the evaluation of others pain: the case of the Jewish-Arab conflict | 2018 | i | 153 | ws | painful events on hands and feet | national/ conflict, neutral | pain intensity rating | y | affective response (sharing), cognitive capacity (understanding) | IEB as an empirical fact | SIT, intergroup conflict, explicit vs implicit priming |
| Physician experience is associated with greater underestimation of patient pain | 2019 | s | 20 | * | real pain | gender | pain intensity rating | y | / | / | similarity (of experiences), gender norms |
| The residential stability mindset increases racial in-group bias in empathy | 2021 | i | 82 | ws | facial expressions of pain | racial | pain intensity and self-unpleasantness rating | n | affective sharing, cognitive PT, cognitive control | IEB as an empirical fact | residential mobility intervention: individual vs collective self-concept, identity salience, social identification and categorization |
| Effects of Patient's Race on Pain Perception and Treatment in Nursing Students (UNPUBLISHED THESIS) | 2021 | s | 104 | bs | painful shoulder movement | racial | pain intensity, severity and attribution | n | / | IEB as an empirical fact | stereotype content and prejudice, salience (of racial issues) |

*Note*. S/I – simple/intervention study concerning IEB; N – sample size; y – yes; n – no;

***Behavioural Studies: Summary***

We undertook this systematic review to give answers to several questions regarding IEB.

The first question was how authors in IEB studies define and measure empathy. For behavioural studies in our review, the traditional division between cognitive and affective qualities of empathy was present in the majority of definitions in the introductory paragraphs. However, there were substantial differences between the study groups. Non-physical pain studies classified as „miscellaneous“ in our review provided the most complex definitions and measured either general empathy or higher-level empathy phenomena such as sympathy, compassion, empathic concern, and perspective-taking (Cameron, 2018). In contrast, misfortunes and physical pain studies, although mostly acknowledging the dual nature of empathy, focused primarily on lower-level emotional components i.e. affect sharing. This focus was evident from the operational definitions of empathy: the former measured congruent affect and the latter pain intensity ratings.

As the main focus of most of the miscellaneous studies which measured more complex outcome variables was not intergroup empathy bias per se, we argue that in behavioural studies IEB was dominantly conceived as an emotional empathy bias. Moreover, we believe this emotional focus was directly inspired by IEB detection in neural resonance studies. We have already illustrated the concept definition of IEB in misfortunes studies. We add several pieces of information to this statement:

1) except for Arroyo (1996) study of general bias in psychotherapists, all studies in our review were published after the Xu et al. (2009) study credited to be the first neuroscience study to find IEB in neural resonance.

2) Moreover, the Drwecki et al. (2011) and the Trawalter et al. (2012) studies, the oldest behavioural studies that entered our review, directly reference neuroscience studies as evidence that empathy is affected by race. This is also the case for the majority of physical pain studies (although some cite an earlier study by Brown et al. 2006, which didn't enter our review).

3) Misfortunes studies also treat IEB as an empirical fact by referencing the Cikara et al. (2014) – the oldest study from our review, which in turn gives the neuroscience studies the credit for "unpacking" the "empathy failures" (Cikara, Bruneau & Saxe, 2011).

In other words, behavioural studies mostly treat IEB as an empirical fact based on dominantly neuroscience data. In addition, some of the respected authorities in the field of neuroscience of empathy explicitly acknowledge the surge of empathy studies mainly because of the improvements in imaging techniques (Coll et al., 2017).

In our opinion, claiming that a phenomenon exists only on the basis of neuroscience data represents an overstatement; in the further text, we will problematise in more detail calling neural resonance an empathic reaction without behavioural validation. Moreover, co-opting the phenomenon from neuroscience without verbalizing the implicit assumptions behind it (for example, that it is automatic and *dampened* or *reduced* for outgroups) and putting it under the same umbrella term "empathy" can mislead the researchers and the public alike that we understand more about it than we actually do.

For example, using pain intensity estimates in physical pain behavioural studies was based on the hypothesis the estimates are affected by the amount of "sharing the other's pain". This assumption is far from proven. Neuroscience studies and behavioural studies do not necessarily measure the same empathic processes. We will discuss this thoroughly in the neuroscience studies section; here it is important to note that in neuroscience studies pain, intensity is hypothesised to represent emotion identification and not emotional sharing; however, pain intensity estimates in the behavioural studies (especially the more complex ones, i.e. using shoulder movement pain or

verbally presented everyday painful events as stimuli) could depend on a myriad of other processes than both emotion identification and emotion sharing (for example, specific stereotypes, see Trawalter et al, 2012).

Another example is Drwecki et al. (2011) citing Avenanti et al. (2010) study as evidence that "individuals experience higher levels of empathic pain for the members of the same racial group" and using this study to hypothesise that empathy bias was the cause of pain treatment bias. However, Avenanti et al. (2010) found bias only in neural resonance but no behavioural bias, citing social desirability as an explanation. Drwecki et al. (2011) was the only pain behavioural study that measured empathic concern, which is a substantially different measure than the difference in intensity of motor evoked potentials.

Misfortunes studies mostly measure congruent affect which is conceptually closer to emotional sharing but again, it is not a proven fact that congruent affect estimates actually rely on neural resonance and could be influenced by a myriad of other factors and circumstances as well (see the discussion on the misfortunes studies below). This is a big problem because we are essentially comparing apples and oranges and putting them into the same fruit basket; however, if you are making a pie, it is very important to know which one it is. In our opinion, just as the everyday label "fruit" is untied to botanical classification, the word "empathy" is ill-defined and if we are speaking of "empathy bias" it is very important to know what exactly we are talking about if we are trying to investigate its causes, consequences and the ways to modify it.

As for the second question – which socio-psychological theories are used to explain the intergroup empathy bias. In total, social identity theory/social categorization theory and its offspring intergroup emotions theory dominated the theoretical rationale. This reflects the vast impact the social identity approach had in social psychology ever since it first appeared. However, it is our impression that the use of these theories and their related concepts in studies whose main topic is empathy is somewhat trivial with respect to studying empathy as a phenomenon. Concepts from the social identity approach were most often used to justify why we expected an intervention to be effective in manipulating intergroup empathy bias (both increasing and reducing it); according to these studies, the fact that intergroup empathy bias existed had already been proven by neuroscience studies. Even when the theoretical rationale was soundly and consistently used to explain the specific hypotheses regarding the "parochial distribution of empathy" (e.g. Bruneau, Cikara & Saxe, 2015), still the main emphasis was on "how empathy influences intergroup conflicts" and the authors accentuate that future studies should examine which outcomes the parochial distribution of empathy predicted best. In contrast, how exactly social identification shapes empathic processes, which ones, and to what purpose – these issues were barely touched in behavioural intergroup empathy bias up to date (see Hasson, Tamir, Brahms, & Cohrs, 2018 for a notable exception).

In our opinion, if we take the social identity approach on empathic reactions such as congruent emotions and empathic concern (especially when studying them in social contexts), to hypothesise that we will feel more of these for the ingroup is almost stating the obvious; however, we don't believe that should necessarily be the case with e.g. emotion identification. In other words, the issue we should explore to gain insight into empathy in intergroup relations is how social identity and various factors influencing intergroup relations shape specific empathic processes in the cascade of empathic reactions and stop muddling them all together by calling them all "empathy". In our opinion, this is currently done so partially because of the conceptual chaos in empathy research as well as implicit assumptions of neuroscience studies were pulled along with the data – authors tend to speak of empathic reactions as automatic, implicit, "natural", and universally present (and we withhold them for the outgroup). However, to us it seems that in an intergroup context (and interpersonal as well), many of the phenomena we list under empathy are actually motivated – employed in certain contexts with certain (implicit or explicit, conscious or unconscious) goals, and thus not universal as claimed.

Finally, according to the behavioural studies we reviewed above, is there an ingroup empathy bias? If we omit the discussion about the reproducibility crisis in social psychology (and we will in this chapter), we can tentatively conclude there is, as the numerical majority of the studies state so. However, the answer becomes very complicated if we decompose it by study groups, i.e. if we problematise the issue of how we elicit and how we measure empathy.[6]

First, studies summarised under the unfortunate label "miscellaneous" were the most heterogenous and most complex with respect to the stimuli aimed to elicit empathic reactions and the theoretical rationale and design within which empathy bias was studied. Half of these studies find ingroup empathy bias, but another half don't, and it seems that finding or not finding bias cannot reliably be attributed to sample size, the ingroup/outgroup division, the task, the dependent variable, etc. Physical pain studies are somewhat heterogenous in the specific manner they present physical pain to participants as well as the general theoretical account, but the nature of the presented events is unambiguous. However, just like emotional pain studies, they sometimes do and sometimes don't find significant bias, which also cannot be reliably attributed to any of the previously mentioned criteria. In addition to presenting relatively unambiguous empathy-eliciting stimuli, misfortunes studies are also very similar to one another in design and rationale. However, in contrast to both emotional and physical pain studies, they typically find intergroup empathy bias.

As the misfortunes studies differ from the former two groups in many ways, there could be a number of reasons for the striking difference in the ratio of "positive" and null findings (17 detecting IEB, 0 with null effects). We point to possible explanations that draw from the social desirability account (i.e. from the hypothesis that people in contemporary egalitarian climate do not readily disclose their biases), but we also provide several alternative explanations that cannot be discounted easily based on the existing data.

Misfortunes studies were dominantly conducted on minimal groups and measured bias as a congruent emotional reaction on a visual analog scale (enabling us to detect very subtle differences) to specific events mostly mild in nature[7]. It could be that this method is particularly suitable for unveiling ingroup empathy bias by cleverly circumventing any social desirability concerns a participant may have (both conscious and unconscious): (1) using VAS makes it difficult to consciously respond exactly the same for ingroup and outgroup targets; (2) displaying bias to mild empathy-eliciting stimuli could be perceived as non-harmful; (3) using minimal groups could easily mask the true goal of the study and loosen or eliminate conscious control of your own behaviour as the context is not "real". However, all the studies were conducted on minimal groups within an explicitly competitive context or groups whose relationships are defined or currently permeated by conflict. Therefore, we cannot be certain if competition or conflict is necessary for empathy biases to emerge in the first place or if it only solidifies, modifies, and/or amplifies them.

In addition, the number of studies of "miscellaneous" studies as well as physical pain studies that do find bias is not negligible, and those include studies conducted on racial groups, where we would expect social desirability concerns to be strongest. All things considered, we don't believe social desirability is an adequate "pan-explanation" for null results without further investigation and that it should be treated as a hypothesis and not a fact.

[6] The goal of this review is to be systematic, i.e. we wanted to explore what exactly the authors mean by "intergroup empathy bias" in empirical studies by performing a formalised search of the key terms. However, it is not a meta-analysis: consistent with our opinion that empathy is undefined and undefinable as a scientific notion, and because behavioural studies are very heterogenous and neuroscientific studies often do not provide enough behavioural data, we believe that calculating effect sizes would be both unfeasible and pointless.

[7] Although some studies include "extreme" events such as "somebody's computer crashed with a week's work on it", we call this events mild when compared to the grand scheme of suffering in the world or to the universally agreed notoriousness of physical pain (although this particular suffering in the world of this paper's author sounds very bad indeed)

However, it could also be the case that there is a reliable empathy bias between "us" and "them" in affect sharing i.e. the intensity of congruent emotional states in the observer, but that bias does not exist by default in perspective-taking or empathic concern (which could in our opinion explain the inconsistent results in both emotional pain studies and physical pain studies in medical context) nor does it exist always in emotion identification (which would explain the inconsistent results in other physical pain studies). We should note that while some authors include emotion identification in the definition of empathy, others treat it as a precursor of empathy but not empathy itself. However, we will further elaborate on this distinction in our discussion of psychophysiological studies, which we turn to next.

## Psychophysiological Studies

We identified 38 reports with 44 studies that measured empathy by registering various signals from the brain and body (but mostly from the brain). The studies were published from 2008 until December 2021. With four exceptions, the publishing pattern was one-article-one-study. All studies were within-subjects with respect to target group identity. The participants were almost exclusively young undergraduates, with very few exceptions (Fourie et al., 2017; Fourie et al., 2019). All studies report ingroup bias in neural activity.

### *Physical Pain Empathy*

In 35 out of 44 studies, participants were exposed to targets experiencing physical pain. A breakdown of studies by painful stimuli is presented in Table 6.

In a typical experiment, participants were presented with painful and neutral stimuli happening to ingroups and outgroups while their brain reactions to those stimuli were being recorded. The most numerous stimuli were images of painful vs neutral facial expressions of ingroup and outgroup faces (14 studies), followed by images or short videos of painful (needle penetration) vs neutral (touch by a q-tip) stimulation performed on neutral faces (7) or hands (8) of ingroups and outgroups (we labelled the latter two "repetitive painful events" as it was always the same stimulation – needle penetration or q-tip touch – on different ingroups' and outgroups' faces). Some studies (2) presented painful and neutral events on the hands and feet of ingroups and outgroups (e.g. knife cuts, paper cuts, a pinky hit on a piece of furniture, etc.), and some (2) presented painful and neutral facial expressions along these variable events. One study presented complex bodily harm happening to different species, humans included.

In some studies (7) participants provided explicit ratings of the stimuli immediately after they were presented, i.e. during the brain scanning. However, in some studies (9) participants were only instructed to watch the stimuli (i.e. had no specific task during scanning), and in the remaining 19 studies participants were performing simple stimuli classification tasks (for example, classifying stimuli by race, painfulness (painful/non-painful) or orientation (upward/inverted)). In the latter two cases, participants were asked to provide their explicit empathy ratings after the scanning procedure (in studies where behavioural ratings were collected at all).

Out of 35 physical pain studies, 4 studies registered no behavioural indicator of ingroup bias, 6 studies registered only reaction times to ingroups and outgroups while performing

classification tasks and/or classification accuracies (which were ceiling high), and the remaining studies (25) asked the participants to provide an explicit evaluation of the empathy-eliciting events.

The mean number of participants is M ≈ 38 and Mdn = 32. Almost 80% of the studies were conducted on racial groups.

1. Ratings. Among studies that required the participants to explicitly evaluate the event and provided information about the ratings in the article[8] (24 or approximately 2/3), 22 asked the participants to rate the pain intensity felt by the target, and 15 asked them to evaluate the self-unpleasantness of the scene.

Table 6
*Type of stimuli by type of study in psychophysiological studies of painful events*

| stimuli | EEG | ERP | fMRI | SCR | tDCS | TMS | total |
|---|---|---|---|---|---|---|---|
| facial expressions of pain | | 10 | 3 | | 1 | | 14 |
| painful events on neutral face (repetitive) | | 2 | 5 | | | | 7 |
| painful events on hands (repetitive) | 2 | | 3 | 2 | | 1 | 8 |
| painful events on limbs (variable) | | 1 | 1 | | | | 2 |
| facial expressions + pain on limbs | | 1 | 1 | | | | 2 |
| painful events: other | | 1 | | 1 | | | 2 |
| total | 2 | 15 | 13 | 3 | 1 | 1 | 35 |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response

Table 7
*Ingroup empathy bias by type of stimuli and type of groups: ratings*

| stimuli | racial | | | non-racial | | | total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Y | N | ? | Y | N | ? | Y | N | ? |
| facial expressions of pain | / | 7* | 6 | 1 | / | / | 1 | 7 | 6 |
| painful events on neutral face (repetitive) | 3 | 2 | 1 | / | 1 | / | 3 | 3 | 1 |
| painful events on hands (repetitive) | 2 | 3 | 2 | / | / | 1 | 2 | 3 | 3 |
| painful events on limbs (variable) | / | 1* | / | / | 1 | / | / | 2 | / |
| facial expressions + pain on limbs | / | / | / | 2 | / | / | 2 | / | / |
| painful events: other | / | / | / | 1 | / | 1 | 1 | / | 1 |
| total | **5** | **13** | **9** | **4** | **2** | **2** | **9** | **15** | **11** |

*Note*. y – bias registered; n – bias not registered? – no ratings; * – outgroup bias

Approximately one in four studies (5 out of 18) on racial groups find ingroup bias in explicit ratings, but two studies find outgroup bias. In contrast, studies on non-racial groups report behavioural bias more often than not (ingroup bias was found in 4 studies and not found in 2 studies). It should be noted though they are less numerous as a category (6 studies vs 18 studies on racial groups).

When presented with facial expressions of pain (8 studies), participants predominantly rated the intensity and unpleasantness of the pain equally for the ingroup and the outgroup (6 studies found no bias and 1 found outgroup bias). The only study finding ingroup bias examined religious groups (the bias is registered in self-unpleasantness ratings).

[8] In one tDCS study, bias was measured as pre-post stimulation ratings difference, i.e. baseline ratings difference of ingroups vs outgroups were not provided in the article.

Studies using repetitive painful events on neutral faces (i.e. needles vs q-tips) as stimuli (6 studies) find ingroup bias in ratings in 50% of the cases. The ratio of studies that find and don't find bias when presenting repetitive painful events on hands is similar (2 studies found ingroup bias and 3 didn't).

Other types of stimuli are used less frequently. The two studies presenting variable (i.e. everyday) painful events on hands and feet, don't find ingroup bias (in fact, one of them finds outgroup bias). On the other hand, both studies that add painful facial expressions to the painful events do report ingroup bias. The study presenting bodily damage to various species that collected empathy ratings finds higher empathy ratings for evolutionary closer animals (and highest for humans).

When analysed by type of rating task, a third of the studies that assessed painfulness find bias (7x ingroup bias and 1x outgroup bias), and the same ratio of studies (5 out of 15) find bias in self-unpleasantness ratings (4x ingroup and 1x outgroup bias). Other types of ratings were used only once but it is interesting that they all indicate ingroup bias.

11 studies in total or somewhat below 50% detected bias in at least one of the behavioural ratings - 9 found ingroup bias and 2 outgroup bias; It is important to remind the reader that only 7 studies collected the ratings during the scanning procedure, and 5 of them find significant differences between the ingroup and the outgroup. In contrast, 17 studies collected the ratings post-scanning, often when presenting the stimuli to the participants for the second time - four found ingroup bias, two found outgroup bias, and 11 found no bias at all.

Table 8
*Ingroup empathy bias by type of stimuli and type of groups: reaction times*

| stimuli | racial | | | non-racial | | |
|---|---|---|---|---|---|---|
| | Y | N | ? | Y | N | ? |
| facial expressions of pain | 4 | 1 | 2 | 2 | | 1 |
| painful events on neutral face (repetitive) | | | | 1 | | 1 |
| painful events on hands (repetitive) | 1 | | | | | |
| painful events on limbs (variable) | | 1* | | 1* | | |
| facial expressions + pain on limbs | | | | | | |
| painful events: other | | | | | 1 | |
| total | 5 | 2 | 3 | 0 | 3 | 0 |

*Note*. x – race classification task; x – pain classification task; x – race and pain classification tasks; x – group classification task; x – painfulness ; Y – bias registered; N – no bias registered; ? –uninterpretable

2. Reaction times. Out of 35 physical pain studies, 15 reported response times as behavioural indicators. In 4 studies, participants were asked to classify the stimuli by race, 5 studies asked the participants to classify the stimuli as painful/non-painful, while 4 studies asked the participants to classify the stimuli by both criteria. Two more studies had no classification requests but reported RT for painfulness ratings provided during scanning. In all studies reaction times during the tasks (i.e. the time needed to respond if a stimulus is painful or not, if the person was Caucasian or Asian, how painful was the stimulus) were compared by group identity (ingroup vs outgroup) and type of stimuli (painful vs neutral).

It is important to note that during race classification, responses were usually faster for neutral than painful stimuli regardless of race, while the reverse is true for pain classification. Therefore, during race classification, IEB would be reflected in especially slow reaction times for ingroup pain (as ingroup pain is expected to disrupt ongoing tasks by its salient nature). However, it is not completely straightforward what should be the pattern of the results that would indicate the existence of IEB during pain classification (nor do any of the authors comment on it). In a study

analysing RTs during painfulness ratings, the authors interpret the longer RTs for outgroup pain as IEB (as we need more time to evaluate how painful is something to a dissimilar other); classifying a stimulus as painful or non-painful is not the same task as rating its painfulness, but if we take this analogy, we should be fastest in classifying ingroup pain.

All studies (4 out of 4) that used only race classification tasks found ingroup bias in RT. However, no study that used both race and pain classification (4 studies) or only pain classification task (3 studies) did (in most cases the results are null, while in some studies it is very difficult to unambiguously interpret the results nor the authors do interpret them). Studies on non-racial groups used pain classification tasks only and all found null effects. As for studies that analysed online RT for painfulness ratings, one does find ingroup bias and one does not.

3. Accuracy. Only 10 out of 35 studies the authors analysed classification accuracy. Two studies, both asking about racial classification, found significant ingroup bias (i.e. they indicate that painful stimuli hinder classification for ingroups but not outgroups, supposing that ingroup pain is a salient stimulus disrupting ongoing cognitive performance). However, in the remaining eight studies there was no difference in classification accuracy. Classification accuracies were very high in all studies and some studies did report high accuracies but did not formally analyse them as it was obvious from raw numbers that the differences did not exist.

4. Brain-behaviour correlations. Eight studies (approximately a fifth of all physical pain studies) reported they had tested the correlations between brain signals and behavioural indicators of empathy. Six out of eight of these studies found bias either in reaction times or behavioural ratings of pain intensity and self-unpleasantness. Two studies that found positive correlations between psychophysiological and behavioural measures interpreted them as meaningful convergence (one conducted on humans vs other animals and the other one on regional groups). In three studies (on racial groups) the correlations were addressed but then ignored in the interpretation no matter the significance, and one study took the absence of correlation as proof that racial bias occurred “at an unconscious level”; interestingly, that is how the remaining two studies out of these eight (both from one report) interpreted the lack of main effects of group on pain and self-unpleasantness ratings, while at the same time, they interpreted the positive correlations between ratings/RT and brain activity as convergent evidence that brain activity really reflected the process hypothesised to indicate. None of the studies addressed the small sample size or very crude behavioural measures (not to mention their validity) as a hypothetical explanation for the lack of correlation.

5. Brain-behaviour discrepancy. As we have already mentioned, all studies report ingroup bias in psychophysiological responses. However, we can unambiguously interpret the behavioural results only in studies in which the participants were asked to directly state their levels of empathy towards ingroups or outgroups. If we take only those into consideration, the researchers observed discrepant results in brain and behavioural indicators in 15 studies: all find ingroup bias in brain signals, but in behavioural ratings some find no behavioural ingroup bias and other find behavoural outgroup bias. Six of these studies either simply note or completely ignore the discrepancy in the behavioural and psychophysical indicators of bias. The nine remaining studies interpret the discrepancy in terms of social desirability or unconscious bias, i.e. explain it by the fact that participants either don’t know or don’t want to show that they are biased.

Among studies that find bias in both brain and behaviour (9), i.e. brain and behavioural measures converge, four studies do not address the correspondence at all (therefore, not addressing the correspondence in studies that found behavioural bias is proportionally equally frequent as not addressing the discrepancy in studies that didn’t). Two studies in this group cite social desirability as important, in this case for finding the bias: one study attributes the fact that the participants were of the same race (and that they were free to state their religious bias), while another interprets different patterns of brain and behaviour bias to various outgroups – both existent – to different

demands social desirability places on brain and behaviour. Two more studies offer ad hoc explanations, while only one study – the one comparing empathic reactions to humans vs other species – interprets the results as validating and explicitly excludes social desirability explanation ("demand characteristics") as unlikely.

6. Definitions of empathy. In 10 out of 35 studies, the authors did not spell out an explicit definition of empathy. The vast majority of the remaining ones (25) define it as "the ability to understand and vicariously share the feelings (and sometimes thoughts) of other people" (17) or differentiates cognitive and affective empathy (4), but some studies limit the definition to emotional sharing (4). A substantial portion of these definitions is followed by an emphasis on automatic and implicit qualities of empathy, its role in motivating prosocial behaviours, and hypothesised sources in shared neural activation. Therefore, despite the explicit acknowledgment of the multifaceted nature of empathy, physical pain studies dominantly deal with affective/experience sharing.

7. Social psychology concepts. In fourteen studies empathy bias was studied or explained by relying on the importance and consequences of social categorization, with six of them referencing the works of Tajfel and Turner, but eight without referring to a specific theory. Some studies investigated the importance of race as a potent social cue and emphasised the role of racial prejudice (2), while some studies (using non-racial groups) dealt with social categorization more generally, by stressing the division of us versus them as important per se (2). A number of studies contrasted the influence of racial vs social categorization within the study design (4) (e.g. by dividing participants into mixed-race minimal groups or adding a third "neutral" group) or examined whether the neural bias is influenced by individuation or recategorization (5) (e.g. by treating the instructions to categorise faces as painful or non-painful instead of by race as "individuation" and minimal group assignment by "recategorization"). We want to draw the readers' attention to the fact that the same interventions were sometimes interpreted through different theoretical notions. For example, while one study considered assigning participants to mixed-race minimal groups as investigating the importance of racial vs. social categorization and treated the intervention as "modifying intergroup relationships", others treated the minimal group assignment as stimuli recategorization. Another nine studies examined whether the bias could be modulated (increased or reduced) by various experimental manipulations or individual difference measures identified (or at least investigated) as important in behaviorual studies of social cognition and general bias: self-construal, mortality salience, residential mobility, physical environment manipulations. Among the last third (12), four studies dominantly relied on the concept of (de)humanization, three dominantly explained their results through dual-route models of cognition, while two studies were focused on the importance of the sensorimotor cortex on social cognition. In addition, three studies dealt with the importance of oxytocin for social cognition, i.e. they did not refer to a specific socio-psychological concept but explored the general effects of oxytocin on social processing in the brain.

All psychophysiological studies examining empathy for physical pain can be found in Table 9.

Table 9

P*sychophysiological studies examining empathy for physical pain*

| Title | Year | Study type | N | Stimuli | Group | Task during scan | Bias in ratings | RT bias | BB correlation tested | Discrepancy | Explanation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Racism and the empathy for pain on our skin | 2011 | SCR | 61 | repetitive painful events on hands | racial (2x) | no task | no ratings | / | / | / | / |
| Racism and the empathy for pain on our skin | 2011 | SCR | 47 | repetitive painful events on hands | racial (2x) | no task | no ratings | / | / | / | / |
| Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses | 2012 | ERP | 16 | painful facial expressions | racial | race categorization, pain categorization | no ratings | no | / | / | / |
| Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses | 2012 | ERP | 16 | painful facial expressions | racial, minimal | race categorization | no ratings | yes | / | / | / |
| Intergroup relationships do not reduce racial bias in empathic neural responses to pain | 2014 | ERP | 21 | repetitive painful events on neutral face | racial, minimal | pain categorization | no ratings | / | / | / | / |
| Mortality salience enhances racial in-group bias in empathic neural responses to others' suffering | 2015 | ERP | 32 | painful facial expressions | racial | pain categorization | no ratings | ? | / | / | / |
| Mortality salience enhances racial in-group bias in empathic neural responses to others' suffering | 2015 | fMRI | 40 | painful facial expressions | racial | pain categorization | no ratings | no | / | / | / |
| Minimal humanity cues induce neural empathic reactions towards non-human entities | 2016 | ERP | 18 | painful events: general | species | pain categorization | no ratings | no | / | / | / |
| Empathic Neural Responses Predict Group Allegiance | 2018 | fMRI | 67 | repetitive painful events on hands | religious | no task | no ratings | / | / | / | / |
| The residential stability mindset increases racial in-group bias in empathy | 2021 | ERP | 38 | painful facial expressions | racial | race categorization, pain categorization | no ratings | / | / | / | / |
| The residential stability mindset increases racial in-group bias in empathy | 2021 | tDCS | 80 | painful facial expressions | racial (2x) | pre-post ratings difference | no info on baseline ratings | / | / | / | / |
| Empathy-related responses to moving film stimuli depicting human and non-human animal targets in negative circumstances | 2008 | SCR | 73 | painful events: general | species | during scan ratings | **general empathy rating** | / | yes | no | validating; SD excluded |
| Do You Feel My Pain? Racial Group Membership Modulates Empathic Neural Responses | 2009 | fMRI | 33 | repetitive painful events on neutral face | racial | pain categorization, post-scan ratings | pain intensity, self-unpleasantness | / | no | yes | unconscious bias |
| Racial Bias Reduces Empathic Sensorimotor Resonance with Other-Race Pain | 2010 | TMS | 36 | repetitive painful events on hands | racial, distant/ unfamiliar | no task, post-scan ratings | sensory and affective pain qualities | / | no | yes | unconscious bias |
| Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses | 2012 | ERP | 16 | painful facial expressions | racial | race categorization, post-scan ratings | pain intensity, self-unpleasantness | yes | yes | yes | unconscious bias, SD |
| Oxytocin modulates the racial bias in neural responses to others' suffering | 2013 | ERP | 16 | painful facial expressions | racial | race categorization, post-scan ratings | pain intensity, self-unpleasantness | yes | no | yes | SD |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response; N – sample size; BB – brain-behaviour; SD – social desirability; **red letters** – ingroup bias; **purple letters** – outgroup bias

Table 9 (continued)

| Title | Year | Study type | N | Stimuli | Group | Task during scan | Bias in ratings | RT bias | BB correlation tested | Discrepancy | Explanation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Their Pain is Not Our Pain: Brain and Autonomic Correlates of Empathic Resonance With the Pain of Same and Different Race Individuals | 2013 | fMRI | 27 | repetitive painful events on hands | racial, distant/ unfamiliar | no task post-scan ratings | pain intensity, pain unpleasantness | / | yes | yes | unconscious bias |
| Shared beliefs enhance shared feelings: Religious/irreligious identifications modulate empathic neural responses | 2014 | ERP | 40 | painful facial expressions | religious | group categorization; post-scan ratings | pain intensity, **self-unpleasantness** | no | no | no | SD |
| Task modulations of racial bias in neural responses to others' suffering | 2014 | fMRI | 21 | painful facial expressions | racial | race categorization, pain categorization, post-scan ratings | pain intensity, self-unpleasantness | no | no | yes | not addressed |
| Challenging emotional prejudice by changing self-concept: priming independent self-construal reduces racial in-group bias in neural responses to other s pain | 2015 | fMRI | 30 | repetitive painful events on neutral face | racial | pain categorization, post-scan ratings | pain intensity, self-unpleasantness | / | no | yes | SD |
| Racial bias in neural response to others' pain is reduced with other-race contact | 2015 | fMRI | 30 | repetitive painful events on neutral face | racial, minimal | during scan ratings | **pain intensity** | / | no | no | no discussion |
| Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity | 2015 | fMRI | 60 | repetitive painful events on neutral face | racial | pain categorization, post-scan ratings | **pain intensity, self-unpleasantness** | / | no | no | sample size, genotype |
| Beta oscillations reveal ethnicity ingroup bias in sensorimotor resonance to pain of others | 2015 | EEG | 69 | repetitive painful events on hands | racial, distant/ unfamiliar | no task, post-scan ratings | **pain intensity,**self-unpleasantness | / | yes | no | no discussion |
| How pain empathy depends on ingroup/outgroup decisions: A functional magnet resonance imaging study | 2015 | fMRI | 30 | variable painful events on limbs | minimal | during scan ratings | pain intensity rating | / | no | yes | not addressed |
| Guess who's coming to dinner: Brain signatures of racially biased and politically correct behaviours | 2016 | fMRI | 25 | repetitive painful events on hands | racial | during scan ratings | pain intensity | yes | no | yes | unconscious bias |
| Degraded perceptual and affective processing of racial out-groups: An electrophysiological approach | 2017 | ERP | 26 | painful facial expressions | racial | scrambled face orientation; post-scan ratings, | pain intensity, self-unpleasantness | / | no | yes | BM inadequacy as an introductory premise |
| Empathy and moral emotions in post-apartheid South Africa: an fMRI investigation | 2017 | fMRI | 38 | painful facial expressions | racial | no task, post-scan ratings | pain intensity | / | no | yes | unconscious bias |
| Physical coldness enhances racial in-group bias in empathy: Electrophysiological evidence | 2018 | ERP | 40 | painful facial expressions | racial | pain categorization, post-scan ratings | pain intensity, self-unpleasantness | ? | no | yes | not addressed |
| Racial Bias in Neural Response for Pain Is Modulated by Minimal Group | 2018 | fMRI | 29 | repetitive painful events on neutral face | racial, minimal | during scan ratings | **pain intensity** | ? | no | no | no discussion |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response; N – sample size; BB – brain-behaviour; SD – social desirability; **red letters** – ingroup bias; **purple letters** – outgroup bias

Table 9 (continued)

| Title | Year | Study type | N | Stimuli | Group | Task during scan | Bias in ratings | RT bias | BB correlation tested | Discrepancy | Explanation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Racial bias in empathy: Do we process dark- and fair-colored hands in pain differently? An EEG study | 2018 | ERP | 23 | variable painful events on limbs | racial | race categorization, pain categorization, post-scan ratings | pain intensity, **self-unpleasantness** | yes (OG) | yes | yes | not addressed |
| Interactions between oxytocin receptor gene and intergroup relationship on empathic neural responses to others' pain | 2019 | ERP | 50 | painful facial expressions | racial | race categorization, post-scan ratings | **pain intensity,** self-unpleasantness | yes | no | yes | not addressed |
| Empathic Responses Are Reduced to Competitive but Not Non-competitive Outgroups | 2019 | fMRI | 69 | variable painful events on limbs, painful facial expressions | rival, neutral | during scan ratings | **pain intensity**, **compassion** | / | no | no | SD influencing brain and behaviour differently |
| Increasing self-other bodily overlap increases sensorimotor resonance to others' pain | 2020 | EEG | 29 | repetitive painful events on hands | racial | no task, post-scan ratings | **pain intensity, self-unpleasantness** | / | yes | no | no discussion |
| The relationship between dispositional self-construal and empathy for ingroup and outgroup members' pain: evidence from ERPs | 2021 | ERP | 27 | repetitive painful events on neutral face | regional | pain categorization, post-scan ratings | pain intensity, self-unpleasantness | no | yes | yes | correlations interpreted as validating, dicrepancy not addressed |
| The neural substrate of schadenfreude: The effects of competition level changes on the processing of pain in others | 2021 | ERP | 40 | variable painful events on limbs, painful facial expressions | political | during scan ratings | **pain intensity**, **schadenfreude** | / | no | no | no discussion |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response; N – sample size; BB – brain-behaviour; SD – social desirability; **red letters** – ingroup bias; **purple letters** – outgroup bias

***Non-physical Empathy.***

Only 9 out of 44 studies (≈ 20%) not using physical pain as an empathy-eliciting stimulus entered our analysis.

In two-thirds of the studies from this group, the researchers used stimuli that could be labelled "emotionally painful". One of these studies did not include any behavioural measure of empathy; in others, participants were asked to rate their empathic concern, personal distress, empathic sadness, or general empathy.

Among the remaining studies, one used Cikara's (mis)fortunes paradigm with congruent emotions rating, while two studies presented participants with emotional (and neutral) facial expressions or action sequences and asked the participants to actively produce them, thus defining empathy as neural correspondence between perception and action. Consequently, the latter studies also did not include any behavioural measure of empathy.

In summary, six studies included behavioural indicators of empathy and four of them found ingroup bias, while two didn't. Interestingly, only one study collected the ratings post-scanning, and that study didn't find ingroup empathy bias (in addition, that is the only study whose direct topic was not empathy and where behaviorual empathy ratings were reported as a secondary measure). The second study that didn't find bias was the misfortunes study, conducted on national groups.

Two studies included non-racial groups: the misfortunes study was conducted on national groups, while the study comparing perception and production of facial expressions used minimal groups. The mean number of participants was M ≈ 55 and Mdn = 32.

Studies that didn't measure empathy behaviourally (3 studies) mainly relied on the PAM (Preston & DeWaal, 2002) and as mentioned, considered neural correspondence to be indicative of empathy or constrained empathy to emotion sharing. The studies that did measure empathy by explicit ratings mention both cognitive and affective qualities of empathy (5 out of 6; one study did not provide an explicit definition of empathy).

Social identity approach-related concepts were again most frequently employed as an explanatory framework (in 5 out of 9 studies). Additionally, Social dominance theory (1) and Stereotype content model (1) were also mentioned, all well as the general influence of prejudice on empathic processes (2).

It is important to note that, in contrast to physical pain studies, most studies (5 out of 6) explicitly tested and interpreted the correlations between brain and behavioural indicators and stressed their importance for validating the psychophysiological signals.

The summary of studies is presented in Table 10.

Table 10

*Psychophysiological studies examining non-physical empathy*

| Title | Year | Study type | N | Stimuli | Group | Task during scan | Bias in ratings | BB correlation tested | Discrepancy |
|---|---|---|---|---|---|---|---|---|---|
| Intergroup differences in the sharing of emotive states: neural evidence of an empathy gap | 2012 | EEG | 26 | facial expressions of sadness | racial | no task | / | / | / |
| Intergroup empathy: Enhanced neural resonance for ingroup facial emotion in a shared neural production-perception network | 2019 | fMRI | 178 | emotional and neutral facial expressions | minimal | no task | / | / | / |
| Perspective taking reduces group biases in sensorimotor resonance | 2020 | EEG | 61 | action sequence | racial | no task | / | / | / |
| Citizens of the World: National Stereotypes Do not Affect Empathic Response in the Presence of Individuating Information | 2021 | fMRI | 21 | negative and neutral events | national | during scan ratings | **congruent emotions (sadness) rating empathic concern, empathic distress, empathic sadness general empathy rating** | yes | no discrepancy |
| Bridging the empathy gap: Effects of brief mindfulness training on helping outgroup members in need | 2017 | EEG | 79 | facial expressions of sadness | racial | no task (ratings post each video, during scanning) | **general empathy rating** | yes | problematizing neural indicators |
| Neural basis of extraordinary empathy and altruistic motivation | 2010 | fMRI | 28 | emotional pain scenes | racial | during scan ratings | **general empathy rating** | yes | validating |
| Cultural influences on neural basis of intergroup empathy | 2011 | fMRI | 27 | emotional pain scenes | racial | during scan ratings | **general empathy rating** | yes | validating |
| Effects of early adversity and social discrimination on empathy for complex mental states: An fMRI investigation | 2019 | fMRI | 36 | emotional distress | racial | no task | personal distress, empathic concern | ? | need for validation recognised |
| Empathy and moral emotions in post-apartheid South Africa: an fMRI investigation | 2017 | fMRI | 38 | emotional distress | racial | during & post scan ratings | personal distress, **empathic concern (post scan)** | yes | unconscious bias, SD |

*Note*. EEG – electroencephalography; ERP – event-related potentials; fMRI – functional magnetic resonance imaging; tDCS – transcranial direct current stimulation; TMS – transcranial magnetic stimulation; SCR – skin conductance response; N – sample size; BB – brain-behaviour; SD – social desirability; **red letters** – ingroup bias; **purple letters** – outgroup bias

### *Psychophysiological Studies: Summary*

A substantial number of studies defined empathy as both cognitive understanding and affective sharing of other's states. However, what they actually measured was brain activation in neural structures or temporal patterns of gross brain activity hypothesised to reflect empathic processing. The majority of the IEB studies investigated empathy for physical pain and either focused on the brain structures active during both reception and observation of physical pain (the so-called "pain matrix") or focused on signals thought to reflect affective sharing. Moreover, half of the non-physical pain studies measured empathy via a brain index of affective sharing of emotional stimuli or defined it as neural congruence between perception and action. In other words, empathy was identified with neural co-activation in the vast majority of studies, and ingroup empathy bias was defined as the differential activation to ingroup and outgroup stimuli.

Our second question was related to the socio-psychological concepts used to explain bias. Although in absolute numbers the concepts related to Social identity theory and Social categorization theory (especially the latter) could be described as numerically dominant, in no study do these concepts play a dominant explanatory role. In other words, the vast majority of studies were neuroscience studies of empathy focused on the underlying neural structures and concepts from social psychology were either used post hoc to explain why the differential activation existed or to design an intervention for mitigating the bias that had been previously proven to exist. Indeed, after the Xu et al., (2009), Avenanti et al. (2010), and Forgiarini et al. (2011) studies that first demonstrated the existence of IEB with fMRI, TMS, and SCR respectively, almost every study included a manipulation aimed to reduce the bias; moreover, the first ERP report of IEB (Sheng & Han, 2012) was designed with the aim to reduce bias.

The third question was, according to the psychophysiological/neuroscience studies reviewed above, is there ingroup empathy bias? The psychophysiological indicators – the differential intensity of brain and bodily responses to ingroup and outgroup painful stimuli – almost unanimously claim there is. On the other hand, behavioural indicators, mostly ratings of pain intensity and/or self-unpleasantness, dominantly show there is not. However, considering the minuscule samples unsuitable for detecting any effect with crude measures such as Likert scales, the fact that the ratio of studies that find and don't find bias varies greatly in different tasks as well as with the procedural differences (during vs post-hoc measurement), and the general disregard for the behavioral measurement as such, we do not stand behind that claim without further investigation.

## General Summary

In keeping with our focus on the elicitation and measurement of intergroup empathy bias, we separately analyzed and summarised behavioural and neuroscientific studies; these were equal in number of studies but not in number of articles, the latter being more numerous. The main findings of this thesis about the concept of intergroup empathy bias in emprirical studies can be summarised as follows:

Behavioural studies that measured differential empathic responses to physically painful events suffered by ingroups and outgroups elicited empathy in several different ways varying in their complexity and ecological validity, but the vast majority operationally defined empathy as pain intensity ratings. Dual nature of empathy (cognitive vs affective) was usually acknowledged in general definitions; however, we argued that both conceptual presentation of intergroup empathy

bias (as an empirical phenomenon relying dominantly on neurosciensce data) and dominant use of pain intensity ratings as indicators of empathy point to an often implicit assumption that affectively resonating with other's pain to a greater extent, in this case with the ingroup (which was demonstrated in neuroscience studies) will lead to higher pain intensity ratings. We pointed out several issues with this assumption: (1) pain intensity ratings are not undisputably considered indicators of empathy, but a precursor to empathic responses; (2) as we move from facial expressions of pain to more complex and contextualised presentations of pain, processes other than affective resonance with other's pain are reasonably expected to contribute to pain intensity ratings. As for the socio-psychological theoretical framework, concepts invited to explain bias ranged from social categorization to stereotypes, but in most cases they were applied to intergroup empathy bias in a rather general fashion.

Among behavioural studies that did not focus exclusively on physical pain, we identified two clusters based on empathy elicitation and measurement.

In the first cluster, labelled misfortunes studies, empathy was typically elicited via descriptions of mildly positive and negative events happening to ingroups and outgroups and operationally defined as congruent affect. General empathy definitions usually stressed its affective nature, and theoretical concepts were used more elaborately and mainly originated from the social identity approach. However, we have illustrated that in these studies the concept of intergroup empathy bias was borrowed from neuroscience studies of empathy for physical pain without explicitly mapping out which ones of the assumptions that go with the data applied for these studies as well.

In the second cluster, labelled miscellaneous, empathy definitions were sometimes absent and when provided, usually acknowledged the dual nature of empathy; empathy-eliciting stimuli were very variable but involved complex negative events, and empathy was operationally defined either as general empathy rating or as average or separate ratings of higher-level responses such as empathic concern, compassion, sympathy. Importantly, only in a minority of studies, the concept of intergroup empathy bias was under investigation; in others, empathy was not the primary topic. Maybe expectedly, the reasoning from the theoretical definition of the concept to tasks and measures in most of these studies is obscured. However, we want to point out that in these studies empathy is sometimes discussed as a motivated phenomenon, in contrast to automatic and universal.

These three groups of studies differ in how frequently they registered ingroup empathy bias. As previously elaborated, the studies vary on many dimensions – groups, tasks, empathy measures, presentation modality, sample type, etc. – and we failed to reach a conclusion as to why this is the case. However, we argued that social desirability is not an adequate explanation to completely account for the divergent study results.

Neuroscience studies of empathy for physical pain are rather unitary in theoretical definitions of empathy, acknowledging the dual nature but stressing its affective qualities and resonant properties. Empathy-elicitation involves presenting simple isolated painful events suffered by the ingroups and outgroups that vary to a degree in ecological validity and contextualisation, but most are either artificial or decontextualised or both. However, operational definitions of empathy are rather homogenous: the primary measure of empathic response is differential activation of neural structures to ingroups' and outgroups' pain. Even though the numerical majority of studies did register behavioural ratings – pain intensity and self-unpleasantness – we argued that they were disregarded in the interpretation of the phenomenon and that the social desirability explanation was drawn as a default explanation. Regarding theories on bias, social identity-related concepts dominate in absolute numbers; however, the articles are better described as atheoretical with respect to the socio-psychological elaboration of findings.

Neuroscience studies that elicit empathy with non-physical events are very heterogenous and comparatively small in numbers. It is worth pointing out that although the primary operational definition of empathy is still neural activation, in this group of studies behavioural indicators of bias appear to be more acknowledged, and efforts to integrate the study results into relevant bias-related concepts from social psychology are sometimes evident.

In summary, it appears that in each group of studies, a specific concept of empathy was investigated, but divergent studies with respect to implicitly hypothesised processes involved in empathic responding were referenced as if they had investigated the same thing. Moreover, we argued that the concept of intergroup empathy bias originated from a collection of empirical findings registering biased neural response to physical pain suffered by ingroups and outgroups. However, these studies were predominantly conducted on racial groups, which prevents us from separating explanations of bias focusing on perceptual expertise and those focusing on social identification. In our opinion, extending the conclusions of these studies investigating a narrow low-level response on small samples in racial groups to a broad, underspecified, and disordered concept such as empathy is extremely premature, especially considering the demonstrated public impact of claims about empathy. In parallel with conceptual development and process specification of empathic phenomena, it is our firm conviction that we first need to establish what are the expected behavioural consequences of the differences registered in our brains.

### *Why Do We Need Behavioural Validation?*

We have already argued for the importance of behavioural validation of neuroscientific measures of psychological processes in general. However, in the context of empathy neuroscience and especially neuroscience of bias we believe behavioural measurement to be critically important, and we offer several arguments for that:

1. Both fMRI and ERP studies in general have been heavily criticised for their analytical flexibility and irreplicable findings (Carp, 2012a; Carp, 2012b; Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafò, 2013; Turner, Paul, Miller, & Barbey, 2018; Paul, Govaart, & Schettino, 2021), and for ERP studies of empathy for pain specifically it has been questioned in recent meta-analyses if the components generally accepted to reflect affective sharing are reliably present in empathy for pain and if they are specific to empathy at all (Coll, 2018). In other words, although all but one study found significant differences in brain activation, our confidence in those results is far from certain.

2. As the number of studies on empathy is on the rise, there have been efforts to introduce conceptual clarity in empathy research and offer a "universal definition"[9], concluding that "empathy as a multidimensional concept with many interacting processes within an overarching *empathic system*" (Guthridge & Giummarra, 2021)[10]. Authors in social neuroscience, while restricting the definition of empathy to adopting another's affecting state, also recognise the paucity of information processing models explaining empathic responses (Coll et al., 2017) which leads to an inability to conclude how exactly a certain factor (for example group membership) exerts its

[9] As we have already stated, our opinion is that empathy is a basket-full-of-everything and value-contaminated to such extent that it is useless in scientific research and should be abandoned as a general term in favour of more specific empathic phenomena.

[10] The meta-definition includes the following dimensions: "Empathy is *the ability* [FUNCTION] to *experience* [PROCESS] *affective* [AFFECT] and *cognitive* states [COGNITION] of *another person*, while maintaining a distinct *self* [SELF-OTHER] in order to *understand* [OUTCOME] the *other*[SELF-OTHER].(Guthridge & Giummarra, 2021, p. 9)"

influence on empathic response. Of particular concern is the differentiation between emotion identification and affect sharing – and differential empathic response in e.g. group settings could stem from group identity influencing the former, the latter, or both. These authors suggest that we need to take into account the changes in pain intensity ratings and use them to validate if the differential empathic brain response should be attributed to emotion identification or affect sharing as well (Coll et al., 2017). As it is clear from our review, a very small minority of studies test even a simple correlation between brain and behaviour, let alone any more complex analysis than that. In our opinion, conducting well-powered behavioural studies using paradigms typically employed in neuroscience can help us identify why is our empathic brain response to outgroups smaller than to ingroups and how we should treat that difference if it does or if it doesn't have empathic behaviorual consequences. In addition, we believe both pieces of information to be useful only when interpreted within a specific model of empathic processing, not in isolation. Whether we do not identify or simply do not share other's pain is a very significant difference in helping us understand both the nature of empathic processes and the nature of bias and inform our decisions on how we should mitigate it or whether trying to do it by targeting individuals is feasible at all.

3. All the reviewed neuroscience studies of empathy bias describe empathy as crucial for social interactions and stress its relationship to prosocial behaviour. The studies following Xu et al. and Avenanti et al. almost universally claim that racial bias in psychophysiological responses can have dire social consequences – *"from medical decisions, rescue operations, police intervention, policy making and, in extreme circumstances, use of physical force and punishment"* (Forgiarini et al., 2011, p. 6)[11] and that we must find a way to reduce *the brain bias* in order to prevent them (e.g. *"discovering methods to reduce RBE-related brain activity has high social importance and may further our understanding of the neurocognitive mechanisms underlying the RBE"*, Sheng & Han, 2012, p. 787). Doing that *"seem(s) important to promote moral progress by overcoming RBE effects and expanding empathy from close persons to humanity in general."* (Fabi & Leuthold, 2018, p 155). At the same time, even the studies that report the correlations between brain and behaviour dominantly neglect it or trivially interpret it, while the brain-behaviour discrepancy is either ignored or easily ascribed to social desirability. The importance of neural resonance is justified by its relationship to behavioural indices in seminal studies (e.g. Singer & Lamm, 2009) (in addition to the sheer importance of empathy for prosocial behaviours); however, we argue that social desirability is not only drawn as a wild card explanation post hoc, but already built-in; however, if the researchers already expect people not to be ready to answer truthfully, it is expected that they value behavioural measures less, and in most cases completely neglect them. Empathy-for-pain paradigms are also called implicit-empathy paradigms (Coll et al., 2017), and brain activity is more frequently correlated to implicit prejudice (measured by IAT in 19 studies and affective priming in 2 studies) than to behavioural empathy measures. Alternative explanations for the lack of behavioural bias, in line with their neglect, are not even noted, let alone seriously discussed.[12] Having that in mind, we find that most of the concluding paragraphs about the importance of IEB in neural indicators could be described as serious overstatements. We believe that neuroscientific

[11] "Our findings suggest that the attitude toward other races may involve not only the overt self-report of the observer concerning attitudes about race but also their deep automatic and physiological reactions. These differential reactions may be elicited even at a very basic level, such as the reaction to physical pain of others. Such a fundamental racial differentiation, in turn, may bias complex activities and judgments over and beyond human consciousness. A precise assessment of other people's pain, in fact, is a necessary skill in many human activities, from medical decisions, rescue operations, police intervention, policy making and, in extreme circumstances, use of physical force and punishment. When all these activities involve people perceived as belonging to different races, a racial bias may hinder pain assessment with detrimental effects on individuals, groups, and their peaceful relationships." " (Forgiarini et al., 2011, p. 6)

[12] Studies that do not use physical pain, in contrast, are much more variable in stimuli, results and different to their relationship to behavioural indicators where they are treated more frequently as validating criteria and in some cases used to problematise neural indicators as such. However, they are much less numerous, confirming the notion that IEB as a term has been inspired and predominantly defined by IEB found in pain studies.

studies of bias threaten to become the "new IAT" and they can be critiqued in a similar manner. Construct validity of implicit association test and its methodological issues have been critiqued by social psychologists, i.e. from within the field (see Schimmack, 2021 and Schimmack, 2021 for further details). However, considering the alleged consequences of intergroup empathy bias, the IAT critique we refer to and find equally important for our discussion comes from outside the field. For example, science journalist Jesse Singal in his book "The Quick Fix: Why Fad Psychology Can't Cure Our Social Ills" (Singal, 2021) includes IAT in the list of several intriguing but half-baked concepts whose public impact had vastly outgrown the evidence behind them, all with good intentions – to make the world a better place. Along with listing the scientific critiques of the IAT, the author stressed that overhyped claims about the power of IAT and the interventions based on those claims led many to believe that the story of racism is simple – that it can be measured and treated on an individual level with simple tools, and if we do that, the progress towards equality was certain – and consequently neglect the complexity of the issue and its structural determinants (Singal, 2021). The word "bias" implies interpretative judgement (Hewstone, Rubin & Willis, 2002); however, currently, there is no firm evidence that IAT scores predict biased behaviour in the real world – hence the use of the term "bias" is not justified except in a very trivial sense (i.e. differences in reaction times) (Singal, 2021). Following the same logic: in empathy research, neural indicators claimed to reflect a process that leads to a certain behaviour are very problematic and without relating them to behavioural measures there is no evidence they do reflect what they claim to. In other words, no effect in the brain has any meaning if it's not related to explicit i.e. measurable behaviour outside the brain. It is, however, dangerous to claim that it does have meaning, as this could lead to essentializing bias among non-experts ("it's in their BRAINS and we MUST edit them").

If the reader believes our worries are exaggerated, please hold that thought: *"Bolstered by recent TMS results suggesting a causal link between mentalizing regions, religious beliefs, and empathic behavior, our present paradigm and classifier may prove useful as an objective diagnostic tool to measure the magnitude of one's ingroup biases (e.g., political party, gender, race). It might therefore prove useful for measuring the efficacy of different interventional programs to reduce the bias between ingroup and outgroup."* (Vaughn et al, 2018, p. 8-9) The narrative is that we would solve the IEB problem if we solve brain bias. This narrative implies several assumptions: (a) that the brain is an entity completely independent from the will of an individual; (b) that we could influence it directly; (c) that we need to influence the brain bias and not what leads to it). None of these assumptions is proven and all are not only scientifically, but philosophically and ethically problematic. However, the concluding passage of a scientific article is usually to a non-scientist the only readable part, and if IAT as a reaction time measure had that much power to shape the public discourse, we believe a brain image to have at least as much. However, what if IEB in neural responses is the inherent response to any kind of "otherness" (e.g. people who like spinach vs people who do not), and we cannot mitigate it or solve it in any way?[13] We cannot know this without process models of empathic responses and knowing our behavioural expectations will help us build them precisely and use the vast potential and the enormous amount of data neuroscience has to offer in the best possible way.

[13] This question is beyond our scope, but a very well powered study that entered our review (Krauterheim et al., 2019) suggests exactly that. They divided the people into two groups based on a fake personality test and asked them to observe and perform emotional and neutral facial expressions of ingroups and outgroups. The number of brain structures with "resonating" properties is larger than expected. This implies that anybody could be the "other" by any criteria and we'd resonate less with him/her if s/he's different from us. In a world where division to us and them is becoming increasingly more complex and abstract, is this finding of any meaning for biased behaviour in everyday circumstances? Does that mean that not only skin colour but personal attitudes and values could also influence our empathy – are the consequences as dire when the line of division is not racial or national? Does that mean that vegan doctor will underestimate the pain of a person wearing a t-shirt from a local bacon festival?

In the empirical part of this thesis, we are going to contribute in several ways, mirroring how empathy was frequently operationally defined in both neuroscience and behavioural studies, in order to attempt to cast light on some of the issues we raised in empathy for pain research.

## Experimental Studies of Physical Pain

### Goals

The first goal of the empirical part of this thesis was to explore in two separate experiments whether behavioural empathy bias could be observed using two types of typical tasks in the neuroscience of pain studies: (Experiment 1) repetitive painful stimuli (needles vs q-tips) (Experiment 2) painful facial expressions and. By presenting these tasks to a substantially larger sample of participants than in typical neuroscience studies while excluding the social desirability explanation, we aimed to provide a benchmark expectation or a frame of reference for other researchers. In other words, we wanted to investigate if we should expect to find ingroup empathy bias at all when using these types of stimuli in socially decontextualised tasks, what is the size of that bias, and if it depended on whether the stimuli intuitively signal pain (painful facial expressions) or the painfulness must be inferred (painful stimuli on neutral faces).To gain insight into specific empathic responses affected by the ingroup-outgroup division, we collected ratings of both pain intensity and self-unpleasantness ratings.

The second goal of this thesis was to investigate ingroup empathy bias for physical pain by using more complex and ecologically valid stimuli. Our focus on the ecological validity of the assessment stems from a firm conviction that if we are drawing implications about intergroup relationships and everyday behaviour of individuals in intergroup contexts, we must examine how the phenomenon in question manifests in situations progressively closer to everyday life and on a broader scope of empathic responses.

In Experiment 3 our goal was to examine whether ingroup empathy bias will emerge in empathic reactions to event-contextualised pain, i.e. everyday painful events not related to group identity happening to ingroup and outgroup individuals.

In Experiment 4, our goal was to examine whether ingroup empathy bias will emerge in empathic reactions to physically painful events happening in social context directly related to the division into „us“ and „them“.

The respondents in these studies were highly identified football fans, a social group known not to hesitate to express their negative feelings towards “the other”, i.e. where social desirability as an explanation could be excluded to a great extent. We tested the fans of two rival teams in the Serbian football league, FC Red Star and FC Partizan, known as “eternal rivals“. We compared the participants’ reactions to expressed and inferred pain of persons wearing the jersey of their own favourite club (the ingroup), their rival (the rival outgroup), and a third club participating in the Superliga (Vojvodina – neutral outgroup in Experiments 1-3 and various neutral outgroups in Experiment 4). By including the third, neutral outgroup as stimuli, we aim to contribute to the explanation of the social mechanism of empathy bias – do we enhance the “default” human empathy for “us”, dampen it for “them”, or both?[14]

Fan groups are particularly suitable for investigating intergroup biases for several reasons:

1) Just like most frequently studied racial groups, they are distinguished by unambiguous visual keys detectable automatically, i.e. it is possible to manipulate group identity of the stimuli without implicating additional processes (compare to the instruction to imagine that the stimulus refers to an ingroup or an outgroup). Compared to racial groups, the advantage of fan groups is that they are arbitrary, their members share important socio-demographic traits (gender, socio-economic

[14] Although formally stimuli wearing Vojvodina’s jerseys could also be counted as “rivals” as they participate in the same league and compete for the same award, no club except Partizan and Red Star has ever won the Serbian league (nor was even close to winning).

status, educational status) and cultural background, and they are not clearly associated with a specific set of stereotypes (e.g. "Asians are smart") that imply systematic differences in other important dimensions (e.g. status).

2) Members of fan groups in the Serbian cultural context are members of the same race, which represents a methodological advantage: it is possible to use completely identical stimuli and manipulate their group identity by varying the colour of the jersey.

3) It is less unacceptable to express animosity towards an outgroup member if they are fan outgroups. This claim is supported by everyday examples of symbolic and physical conflicts between football fans, as well as by a separate research line investigating the related phenomenon of intergroup Schadenfreude – these studies suggest that fans readily report joy to events such as an outgroup player suffering concussion, even when there is no material benefit for the ingroup (Schurtz, Combs, Hoogland, & Smith, 2014).

In short, fan groups are similar to minimal in their arbitrariness, but in contrast to minimal groups they are "real" and psychologically relevant, with clear and easily manipulable visual identifiers. In addition, their relationship is straightforward and social norms about displaying bias in this relationship are loose, which makes them suitable for drawing conclusions about important theoretical concepts in the study of bias.

Additionally, we collected measures of empathy dispositions and implicit bias frequently used in neuroscience studies to gain insight into how they are related to our explicit assessment of ingroups and outgroups.

As empathic brain bias to painful stimuli was more frequently correlated to implicit prejudice than behavioural ratings (with some studies finding a positive relationship between IEB and IAT scores, e.g. Fabi & Leuthold, 2018, and some finding null effects, e.g. Sheng et al., 2014) consistent with the expectation that expressing bias was subject to social desirability concerns, we wanted to investigate how implicit bias was related to behavioural ratings when such concerns were minimised.

As for trait empathy, in some behavioural studies of intergroup empathy bias (e.g. Hanson, 2017), as well as in Bloom's widely discussed critique of empathy (Bloom, 2016), trait empathy was seen as parochial as it increases our empathic response for our cause – whatever it may be. It follows that people who see themselves as more empathic would be more prone to parochial distribution of empathy, which was shown in some neuroscience studies (e.g. Li et al., 2015). Other neuroscience studies point to the positive relationship between trait empathy and the size of the brain empathic reactions regardless of ingroup group biases, and found that more trait-empathic people are more responsive to interventions aimed to reduce brain IEB (e.g. Sheng & Han, 2012); this speaks in favour of the opposite hypothesis – that more empathic people would be less prone to empathy parochialism. Other studies (e.g. Bruneau, Cikara & Saxe, 2017), contrast the contribution of trait empathy vs parochial distribution of empathy (differential empathy for us and them) to predicting outgroup-related outcomes, i.e. see the trait empathy and intergroup empathy bias as independent. We added the measure of trait empathy to investigate its relationship with both the size of the empathic reactions to the ingroups and outgroups and the difference between them, i.e. intergroup empathy bias. In other words, we wanted to contribute to the discussion on if and how trait empathy was related to intergroup empathy bias when assessing physical pain.

Additionally, we included a measure of Social dominance orientation (Pratto, Sidanius, Stallworth, & Malle, 1994), a trait measure of preference for inequality among groups, with exploratory goals. As inequality among groups is not only reflected but created by biased perception, affect, and behaviour, preference for inequality among groups as a trait is expected to be related to various forms of bias; that has been reliably demonstrated in empirical studies of bias (e.g. Levin, Federico, Sidanius, & Rabinowitz, 2002), and examined in some studies of biased empathic responses with inconclusive results (e.g. Cheon et al., 2011; Grimm, 2016).

The empirical part of this thesis will be presented in two separate sections:

1. In *Physical Pain in the Laboratory* we present and discuss the experiments using the typical pain empathy paradigms in neuroscience studies.

2. In *Physical Pain in Context* we present and discuss the results of the two experiments using more complex and ecologically valid painful stimuli to test the existence of intergroup empathy bias.

As all studies share several design features, we will first describe the general procedure and common elements in all four experiments. All studies were preregistered (aspredicted.org; #118778, #118782, #40095). All study materials (including stimuli and raw data for all studies) are available on OSF (https://osf.io/gd5c8/).

## General Outline

### *Participants and Procedure*

Participants were recruited via various channels. A portion were students of three faculties in the University of Belgrade (Faculty of Sport and Physical Education, Faculty of Organizational Sciences, and Faculty of Mechanical Engineering), who completed the study for course credit, and volunteers recruited from various online sources. First, the link to the study was shared on social networks (Facebook profiles and groups, Twitter, Instagram), and in comments below online sports news. Moreover, it was also distributed in-person on live sports events, as well as in relevant football discussion topics on various Serbian forums (Reddit Serbia, forum.b92.net, parapsihopatologija.com, vasudeva.forumburundi.com, mycity.rs). In addition, a portion of the sample was collected by snowball sampling, by recruiting football fans among authors' close and weak ties to distribute the link among their peers.

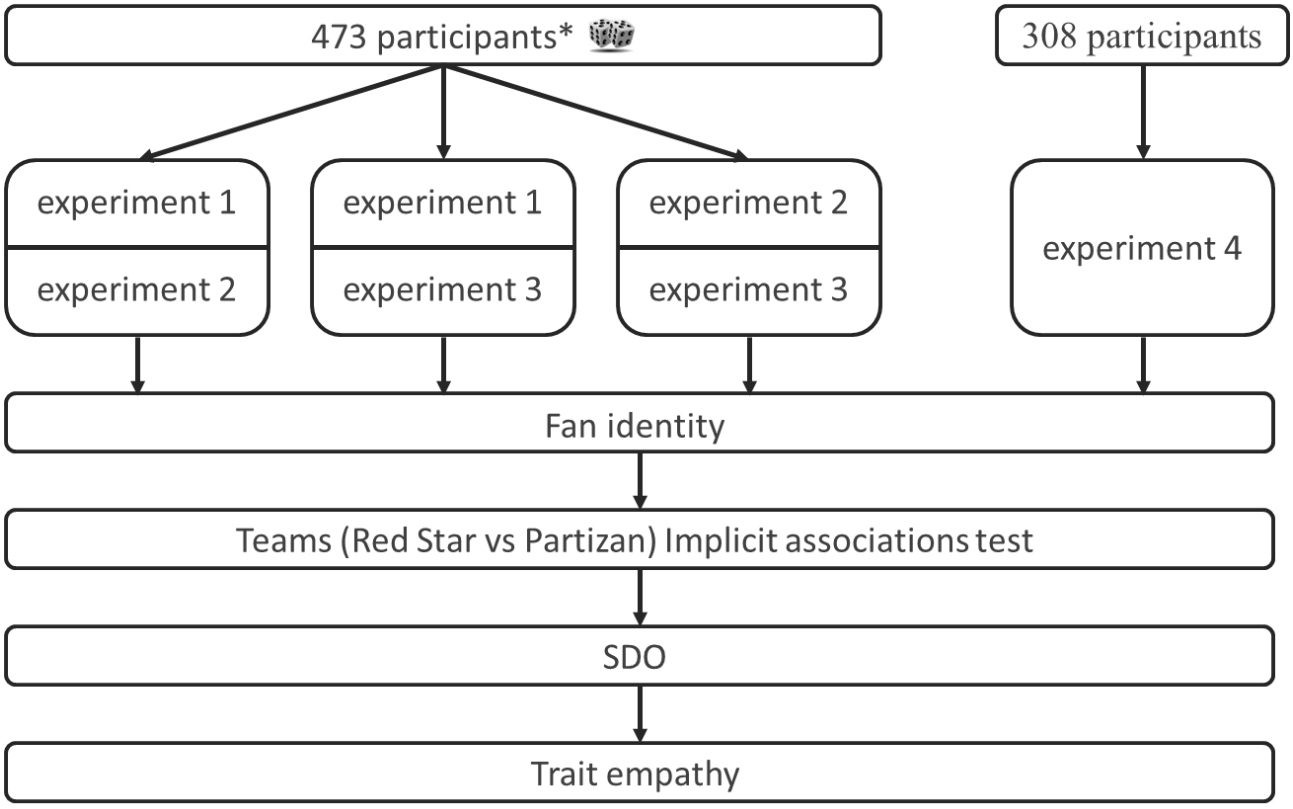
The studies were conducted online via soscisurvey.de, through two separate procedures (i.e. via two separate links). A general outline of the studies is presented in the flowchart below.

After reading the study description, consenting to participate in the study and providing socio-demographic data, participants first completed the experimental tasks, followed by fan identity measures, the Teams Implicit association test (IAT) (Red Star vs Partizan), and then self-report questionnaires.

Taking part in the study was voluntary and anonymous. On the last page, participants were debriefed and thanked; students participating for course credit were redirected to a separate questionnaire where they provided their student id to be compensated for participation. The completion of the study for both Experiments 1-3 and Experiment 4 took ≈ 30 minutes to complete.

Figure 2

*Experimental flowchart*




Participants completed the tasks and questionnaires in the order they are described below.

Socio-demographic information

Participants were asked to state their gender, age, the highest level of education obtained, student status, and employment status, as well as to assess their socioeconomic status on a visual slider as compared to their reference group (i.e. other people living in Serbia). Socioeconomic status was described to them as economic, educational, and vocational status combined.

1. Experimental task

As mentioned, data for Experiments 1-3 and Experiment 4 were collected separately.

In Experiment 1 participants were presented with painful and neutral events happening to ingroups, rival outgroups, and neutral outgroups and asked to rate their empathic reactions to each stimulus.

In Experiment 2 participants were presented with painful and neutral facial expressions of ingroups, rival outgroups, and neutral outgroups and asked to rate their empathic reactions to each stimulus.

In Experiment 3 participants watched videos of persons – ingroups, rival outgroups, and neutral outgroups – describing a painful event that had happened to them. The event could be either identity-related i.e. salient or identity-neutral i.e. non-salient. Participants were asked to rate their empathic reactions to each stimulus.

In Experiment 4 participants were asked to rate their empathic reactions to images of real football fouls in which both the victim and the perpetrator could be a member of the ingroup, the rival outgroup, or the neutral outgroup (excluding the impossible combinations such as the ingroup being both perpetrator and victim).

In all experiments, the ingroup-outgroup status of the stimulus depended on the identity of the participant. For Red Star fans, models wearing a Red Star jersey or Red Star football players represented the ingroup, those wearing a Partizan jersey the rival outgroup, and vice versa for Partizan fans. For both fan groups, stimuli with models wearing Vojvodina or other clubs' jerseys were counted as neutral outgroups.

For Experiments 1-3 participants (N = 472) were randomly allocated to complete two pain empathy tasks: tasks for Experiments 1+2, 1+3, or 2+3. Within each combination, there were three versions of each task in Experiments 1 and 2, depending on the jersey the model was wearing (Red Star, Partizan, and Vojvodina), with different model-jersey combinations presented to different participants. For example, model 1 was presented as a Red Star fan to one set of participants, a Partizan fan to the second set, and a Vojvodina fan to the third set of participants. In addition, in Experiment 3 the model retells two versions of the same event (salient and non-salient) in each jersey, which meant there were 6 model-jersey-event combinations. Each participant completed two experimental tasks and was presented with a consistent model-jersey combination (same in both tasks). Model-jersey combinations were arranged in Latin square block design to account for any possible effects of individual faces and/or posing/acting skills on participants' ratings.

2) Data from 308 participants was collected for Experiment 4. In experiment 4, all participants were presented with the same experimental task, described in more detail in the corresponding section.

An illustration of the stimuli sets is presented in Figure 3. Stimuli design and pretesting for Experiments 1-3 are described in more detail in Appendix C.

After completing the experimental tasks (1-3) or task (4), participants were forwarded to a set of questions about their fan identity.

## 2. Football fan identity

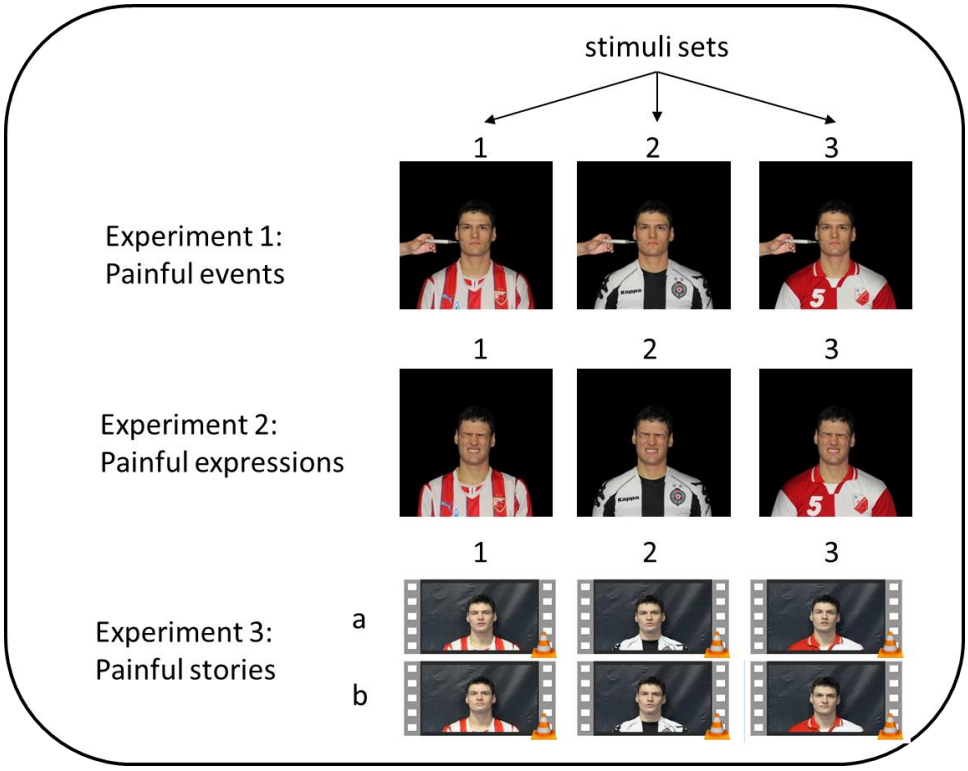
In this block, all participants indicated whether they watched football (yes/no) and whether they supported a specific club (yes/no). Participants who stated they do support a club provided further information about their fan identity. We used the Sport spectator identification scale (Wann and Branscombe, 1993) to measure the strength of identification with the team, which consisted of 7 items followed by an 8-point Likert scale. However, we only included the first five out of seven items on SSIS in the average score, as the sixth and seventh items are (1) substantially different in content than the rest of the scale, (2) have significantly lower averages than the rest of the scale, and (3) by excluding them the reliability of the scale increases (Appendix D).

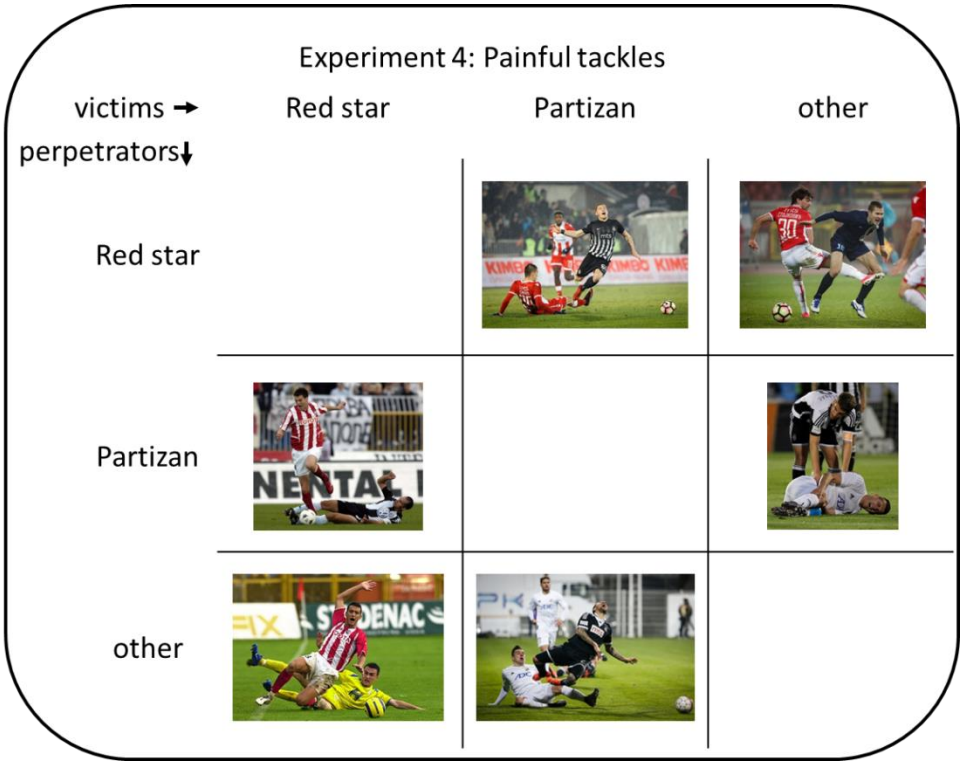
In addition, we used BIRGing (Basking In Reflected Glory, i.e. fan behaviours following their team's victory) and CORFing (Cutting Off Reflected Failure, i.e. fan behaviours following their team's defeat) scales (Spinda, 2011) to assess impression management behaviour related to sports fandom. Both scales consisted of 6 items, followed by a 5-point Likert scale. Both were back-translated into Serbian for this study. Additionally, we asked the fans to rate their team's

status as compared to their rivals, their similarities, and differences, as well as to provide additional information about their virtual and live match attendance behaviour.

Figure 3

*Stimuli sets in Experiments 1-4*



3. Implicit association test

The following block consisted of an implicit association test (IAT), which measures the strength of associations between target categories (in this case, Red Star and Partizan football clubs) and two categories of attributes (good/bad, pleasant/unpleasant). In this study, we used the complete IAT with seven blocks, as detailed in the table below. In the Red Star-Partizan classification task, target categories were represented with 10 unambiguous images of everyday objects with club insignia and club-specific colours (crests, jerseys, scarves, flags, and hats). In the evaluative decision task, we used five positive and five negative verbal stimuli representing evaluative categories good (victory, love, beauty, happiness, laughter) and bad (defeat, hatred, ugliness, sorrow, cry). The order of the stimuli was pre-randomised in each block. The D measure was calculated by using the algorithm recommended by Greenwald, Nosek & Banaji (2003). Latencies < 300 ms and > 10000 ms, and participants with more than 10% of latencies <300 ms were removed, error latencies were replaced by mean correct latencies + 600 ms, and pooled SD was calculated using all latencies (both correct and error). In addition, both practice (3 and 6) and test double classification blocks (4 and 7) were used for calculating the D measure. D measure was calculated separately for practice and test blocks and represents the difference between the average latencies of those blocks divided by their pooled SD. Those two D measures were then averaged to get the IAT D score. Scores above 0 point to an implicit preference for Red Star, and scores below 0 to an implicit preference for Partizan.

Table 11

*Red Star vs Partizan implicit associations test*

| block | classification task | N of trials | stimuli | left key | right key |
|---|---|---|---|---|---|
| 1 P | single | 20 | images | Red Star | Partizan |
| 2 P | single | 20 | words | good | bad |
| 3 P | double | 20 | images and words | Red Star OR good | Partizan OR bad |
| 4 T | double | 40 | images and words | Red Star OR good | Partizan OR bad |
| 5 P | single | 40 | images | Partizan | Red Star |
| 6 P | double | 20 | images and words | Partizan OR good | Red Star OR bad |
| 7 T | double | 40 | images and words | Partizan OR good | Red Star ORbad |

*Note*. P – practice, T – test.

4. Interpersonal reactivity index

The Interpersonal reactivity index (IRI; Davis, 1980) was used for measuring dispositional empathy defined as “a reaction of an individual to the observed experience of another”. IRI consists of 28 items divided into four subscales: Perspective (tendency to spontaneously adopt another person’s viewpoint), Fantasy (tendency to transpose oneself into feelings and actions of fictitious characters), Empathic concern (tendency to feel compassion and care towards the less fortunate) and Personal distress (tendency to feel anxious and discomfort in tense interpersonal situations). Each item is followed by a 5-point Likert scale ranging from *does not describe me* to *describes me very well*, and scores are calculated by averaging the responses for each subscale separately after reverse-scoring the negatively phrased items.

5. Social dominance orientation

$SDO_7$ scale (Ho, Sidanius, Kteily, Sheehy-Skeffington, Pratto, Henkel, Foels, & Stewart, 2015) was used to measure orientation towards social dominance as a personal disposition defined as preference towards hierarchical relationships between groups, i.e. preference for inequality. It consists of 16 items phrased as statements for and against a kind of ideal or desired state in intergroup relationships related to dominance and anti-egalitarianism. Each item is followed by a 7-point Likert scale, and the total score is calculated by averaging the responses to *for* and (reverse-scored) *against* statements.

### *Inclusion and Exclusion Criteria*

In this step, participants were excluded per dataset, i.e. separately for Experiments 1-3 and Experiment 4. However, the same inclusion criteria were applied to both datasets, and participants were excluded if not meeting any of the following conditions:

1. gender: they were not male (1-3: N = 50; 4: N = 52)

2. fan identity: they did not identify as fans (1-3: N = 156; 4: N = 92) or they identified as fans of clubs other than Red Star/Partizan (1-3: N = 15; 4: N = 14)

3.response quality: their responses were "too fast, too straight, or too weird", i.e. their relative speed index was above the recommended threshold. Relative speed index (TIME_RSI) is a soscisurvey built-in non-reactive measure of careless responding (Leiner, 2019). We excluded all participants whose TIME_RSI value was above 2.0, which is the recommended critical threshold (1-3: N = 33; 4: N = 5).

In addition, 17 participants were excluded from the Experiments 1-3 dataset because they recognised at least one of the models from the stimuli set, and 1 participant was excluded from Experiment 4 dataset because he was under 15, which is below the age of consent according to Ethical guidelines of Serbian Psychological Association.

After applying the listed criteria[15], 257 participants in the Experiments 1-3 dataset and 179 participants in the Experiment 4 dataset remained.

In our cultural context, a lot of people will readily align with Red Star or Partizan not because they truly value their fan identity, but because the club in question is symbolically associated with their important others (family, friends, partners). To ensure that only "true fans" entered our sample, we further excluded participants whose average score on the Sport spectator identification scale (SSIS; Wann & Branscombe, 1993; Bernache-Assollant, Bouchet, & Lacassagne, 2007) was below the theoretical midpoint (i.e. below 4).

Applying the last criteria left us with 204 participants in Experiment 1-3 dataset (Experiment 1: 147; Experiment 2: 140; Experiment 3: 121) and 158 participants in the Experiment 4 dataset. In Experiment 4 dataset there were no missing data. In the Experiment 1-3 dataset five entries for IRI and SDO scales were missing (N = 199). In both dataset, we deleted the teams IAT scores for participants with > 10% very fast responses, as per previously described IAT scoring procedure (N = 13 in Experiments 1-3 and N = 18 participants in Experiment 4 dataset).

[15] Some of these participants fulfilled more than one exclusion criteria.

## Experiment 1 and Experiment 2: Physical Pain in the Laboratory

In Experiment 1 and Experiment 2, we presented the participants with two sets of stimuli frequently used in neuroscience studies that aim to elicit empathy for physical pain:

1. Experiment 1: painful vs neutral events on neutral facial expressions

2. Experiment 2: painful vs neutral facial expressions,

and asked the participants to rate the pain intensity and self-unpleasantness for each stimulus.

The faces in each task belonged either to the ingroup, to the rival fan outgroup, or to the neutral outgroup. A differential reaction to painful and neutral stimuli (i.e. the difference in ratings between pain and neutral) would indicate that the participant responded with empathic reactions to painful events and expressions.

Studies up to date using the same tasks dominantly but not unanimously show no ingroup empathy bias in explicit ratings. As discussed in previous chapter, we believe further studies are needed to confirm or disprove this claim.

If after recruiting a larger sample and minimizing social desirability concerns no differences in ratings were found, this would support the conslusion that intergroup empathy bias is absent in explicit ratings.

Alternatively, if we detected differences in empathic responding between the ingroup and the outgroups, three scenarios are possible:

(a) higher empathy for the ingroup compared to both outgroups would suggest ingroup love as the mechanism of intergroup empathy bias;

(b) higher empathy for the ingroup compared only to the rival outgroup would suggest outgroup hate as the mechanism of creating intergroup empathy bias

(c) ratings ordered ingroup – neutral outgroup – rival outgroup would suggest that both ingroup love and outgroup hate were at stake.

Additionally, we expected the effect, if observed, to be stronger in stimuli presenting painful vs neutral facial expressions (Experiment 2) than painful vs neutral events (Experiment 1), as the former directly communicates pain and the latter implies pain but does not signal pain directly.

### Method

#### *Participants*

Out of 204 participants in this dataset, 64 participants completed only Experiment 1, 57 participants only Experiment 2, and 83 participants completed both experiments.

In the total dataset, 127 (62.3%) participants identified themselves as Red Star fans, and 77 (37.7%) of the participants identified themselves as Partizan fans. Participant's mean age was somewhat below 25 years; however, the median age of our sample was below the mean ($M_{age}$ = 24.6 vs $Mdn_{age}$ = 21 vs $Mod_{age}$ = 20), and the majority of participants (69.6%) marked their status as "currently studying". Our participants rated their socioeconomic status somewhat above average (M

= 58.68 on a 1-101 unmarked slider scale anchored at *people in the worst situation* vs *people in the best situation*).

Despite our efforts to diversify the sample, these data suggest our sample dominantly consisted of young university students; however, the number of non-students was not negligible and a quarter of the sample was older than 25 ($Q_{75}$ = 24.75), indicating we managed to a certain degree to capture a wider population of fans. This was especially the case with Red Star fans, who tended to be older ($M_{age}$ = 25.6, $Q_{75}$ = 28 vs Partizan $M_{age}$ = 22.8, $Q_{75}$ = 23) and had a smaller percentage of students (66.1% in Red Star vs 75.3% in Partizan fans).

Experiment 1 was completed by 147 participants and Experiment 2 by 140 participants. No notable differences between these two groups were found.

A more detailed description of the total sample, as well as sample by fan identity and sample by experiment, is presented in Appendix E.

### *Stimuli and Experimental Tasks*

In both tasks, 12 different models were presented to participants. Participants were told that the photographs come from a previous researcher's project on football fans and pain and were instructed to provide ratings after each image.

Four models were wearing Red Star jerseys, four were wearing Partizan jerseys and four were wearing Vojvodina jerseys. Ingroup-outgroup outgroup status of the stimulus depended on the participants' fan identity (Red Star or Partizan), and models wearing Vojvodina jerseys counted as the neutral outgroups for both fan groups.

1.1 In Experiment 1, models receiving painful (needles) or neutral (q-tips) stimulation on their left and right cheeks (12 models x 2 painful/neutral x 2 left/right cheek = 48 trials in total) were presented to participants. Models' jerseys signaled if they belonged to the ingroup, the rival outgroup, or the neutral outgroup, making this a 3(group: ingroup, rival outgroup, neutral outgroup)x2(event: painful, neutral) within-subjects study.

1.2 In Experiment 2, models presenting painful and neutral facial expressions were presented to participants (12 models x 2 painful/neutral = 24 photos). Models could belong to the ingroup, the rival outgroup, or the neutral outgroup, making this a 3(group: ingroup, rival outgroup, neutral outgroup)x2(event: painful, neutral) within-subjects study.

The presentation order of the stimuli was randomised for each participant. In both experiments, participants were asked to rate the sequentially presented images for the intensity of the painful event/expression in the image (pain intensity: *not at all intense – very intense*) and how unpleasant they felt while watching the image (self-unpleasantness: *not at all unpleasant – very unpleasant*) of each image on a 1-7 Likert scale. There was no time restriction for stimulus presentation and for providing ratings.

## Results

### *Intergroup Empathy Bias: Main Analyses*

There were 6 ratings per outcome measure (pain intensity and self-unpleasantness) in each experiment. Descriptive parameters for pain intensity and self-unpleasantness ratings of painful and neutral events happening to ingroups, rival outgroups, and neutral outgroups (Experiment 1) and painful and neutral facial expressions of ingroups, rival outgroups, and neutral outgroups (Experiment 2) are presented in Tables 12 and 13.

For painful events, mean ratings of both pain intensity and self-unpleasantness are of similar intensity and group on the low end of the 1-7 scale, indicating the participants didn't consider the images of painful events, particularly intense or unpleasant. For neutral events, both outcome ratings for all groups fall somewhere between 1 and 2, thereby confirming their neutrality.

Although in all variables there were scores in the upper range of the scale, most of the scores group in the lower range, as also indicated by the median which is lower than the mean in all cases and close to the value of the third quartile. In addition, in this experiment, all scores deviate from the normal distribution as indicated by the Shapiro-Wilks test. For painful events, both ratings for all groups are positively skewed, indicating a smaller set of participants tended to give somewhat higher ratings. For neutral events, both ratings are heavily positively skewed and leptokurtic (for zSk and zKu and additional descriptives of scores and distributions see Appendix F), indicating most participants rated the events as not at all painful or unpleasant.

Table 12

*Experiment 1: Pain intensity and self-unpleasantness ratings for ingroups, rival outgroups, and neutral outgroups for painful and neutral events*

| | pain intensity rating | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | painful | | | | neutral | | | |
| | M(SD) | Mdn(Q75) | Range | SW | M(SD) | Mdn (Q75) | Range | SW |
| ingroup | 2.66 (1.50) | 2.25 (3.75) | 1-7 | 0.905** | 1.60 (0.92) | 1.13 (2.00) | 1-6.25 | 0.693** |
| rival outgroup | 2.72 (1.52) | 2.38 (3.63) | 1-7 | 0.905** | 1.65 (0.97) | 1.25 (2.00) | 1-7 | 0.707** |
| neutral outgroup | 2.65 (1.49) | 2.25 (3.38) | 1-6.13 | 0.899** | 1.57 (0.86) | 1.25 (1.88) | 1-5.50 | 0.709** |
| | self-unpleasantness rating | | | | | | | |
| | painful | | | | neutral | | | |
| ingroup | 2.53 (1.65) | 2.00 (3.50) | 1-7 | 0.850** | 1.53 (0.95) | 1 (1.75) | 1-6.25 | 0.630** |
| rival outgroup | 2.72 (1.83) | 2.00 (3.75) | 1-7 | 0.846** | 1.74 (1.31) | 1 (2.00) | 1-7 | 0.626** |
| neutral outgroup | 2.62 (1.75) | 2.00 (3.50) | 1-7 | 0.843** | 1.56 (1.04) | 1 (1.75) | 1-6.25 | 0.607** |

*Note*. M = Mean; Mdn = Median; Q75 – value of the 75th percentile; SW – Shapiro-Wilks statistic; * p < .05, ** p < .01

Table 13

*Experiment 2: Pain intensity and self-unpleasantness ratings for ingroups, rival outgroups, and neutral outgroups for painful and neutral facial expressions*

| | pain intensity rating | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | painful facial expression | | | | neutral facial expression | | | |
| | M(SD) | Mdn (Q75) | Range | SW | M(SD) | Mdn Mdn (Q75) | Range | SW |
| ingroup | 4.11 (1.24) | 4.25 (5.00) | 1-6.75 | 0.983 | 1.42 (0.64) | 1 (1.50) | 1-4.50 | 0.703** |
| rival outgroup | 4.25 (1.36) | 4.50 (5.25) | 1-7 | 0.977* | 1.61 (1.01) | 1.25 (1.75) | 1-7.00 | 0.657** |
| neutral outgroup | 4.18 (1.26) | 4.25 (5.25) | 1-6.75 | 0.979* | 1.56 (0.88) | 1.25 (1.75) | 1-5.75 | 0.683** |
| | self-unpleasantness rating | | | | | | | |
| | painful facial expression | | | | neutral facial expression | | | |
| ingroup | 2.61 (1.58) | 2.25 (4.00) | 1-6.50 | 0.882** | 1.33 (0.61) | 1 (1.25) | 1-4.25 | 0.603** |
| rival outgroup | 2.84 (1.85) | 2.25 (4.25) | 1-7.00 | 0.868** | 1.71 (1.39) | 1 (1.75) | 1-7.00 | 0.578** |
| neutral outgroup | 2.76 (1.74) | 2.00 (4.00) | 1-6.75 | 0.869** | 1.50 (0.93) | 1 (1.50) | 1-6.50 | 0.613** |

*Note*. M = Mean; Mdn = Median; SW – Shapiro-Wilks statistic; * $p < .05$, ** $p < .01$

For painful expressions, mean ratings of pain intensity were somewhat above the theoretical midpoint of the scale, with the median being higher than the mean. Mean ratings of self-unpleasantness are lower, with the mean higher than the median and they grouped on the low end of the 1-7 scale. This indicates our participants on average considered the painful expressions to be of medium intensity but did not consider them particularly unpleasant. For neutral events, like in Experiment 1, both outcome measures fall somewhere between 1 and 2, thereby confirming their neutrality.

For painful expressions, pain intensity scores were symmetric and mesokurtic (Appendix F) and did not deviate from the normal distribution as indicated by the Shapiro-Wilks test at the $p < .01$ threshold. However, unpleasantness ratings for all groups were positively skewed, as in Experiment 1, and deviated from the normal distribution. The results were similar to Experiment 1 for neutral events too: both ratings were heavily positively skewed and leptokurtic (Appendix F), and scores grouped at the lower range (median lower than the mean and close to the third quartile) and all scores deviated from the normal distribution as indicated by the Shapiro-Wilks test.

Pain intensity ratings for painful stimuli were notably higher in Experiment 2 where participants rated painful facial expressions as compared to painful events on neutral facial expressions in Experiment 1. However, self-unpleasantness ratings, as well as both ratings for neutral stimuli were remarkably similar.

As ratings for neutral events were heavily skewed, we decided to present Spearman's Rho correlations between intensity and unpleasantness ratings for ingroups, rival outgroups, and neutral outgroups within and between experiments (Table 14). As a subset of participants completed both Experiment 1 and Experiment 2, the diagonal contains the correlations of outcome measures across Experiments (N = 83). Intercorrelations of outcome measures within Experiment 1 (N = 147) are shown below the diagonal, and intercorrelations of outcome measures within Experiment 2 (N = 140) are shown above the diagonal.

Table 14

*Experiment 1 & 2: Correlations between ratings*

| | | | pain intensity | | | | | | self-unpleasantness | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | painful | | | neutral | | | painful | | | neutral | | |
| | | | IG | OG | DG | IG | OG | DG | IG | OG | DG | IG | OG | DG |
| pain intensity | painful | IG | **0.073** | .659** | .779** | .293** | .272** | .356** | .497** | .301** | .407** | 0.051 | 0.125 | .188* |
| | | OG | .911** | **0.145** | .763** | .182* | .376** | .380** | .396** | .549** | .493** | 0.064 | .314** | .303** |
| | | DG | .946** | .933** | **0.102** | .332** | .324** | .401** | .510** | .437** | .545** | 0.130 | .297** | .322** |
| | neutral | IG | .431** | .406** | .423** | **.322**** | .649** | .593** | .284** | .184* | .262** | .584** | .407** | .474** |
| | | OG | .352** | .435** | .387** | .843** | **.453**** | .652** | .273** | .387** | .363** | .395** | .561** | .524** |
| | | DG | .345** | .343** | .379** | .864** | .844** | **.331**** | .288** | .296** | .352** | .458** | .497** | .640** |
| self-unpleasantness | painful | IG | .634** | .574** | .610** | .256** | .213** | .206* | **.310**** | .800** | .879** | .425** | .471** | .523** |
| | | OG | .527** | .599** | .556** | .199* | .273** | .166* | .864** | **.394**** | .899** | .382** | .647** | .621** |
| | | DG | .575** | .577** | .622** | .237** | .230** | .222** | .912** | .927** | **.344**** | .408** | .643** | .635** |
| | neutral | IG | .254** | .248** | .262** | .619** | .592** | .572** | .477** | .445** | .452** | **.604**** | .535** | .595** |
| | | OG | .190* | .253** | .225** | .507** | .653** | .533** | .410** | .570** | .475** | .805** | **.607**** | .805** |
| | | DG | .254** | .256** | .293** | .560** | .583** | .658** | .445** | .479** | .500** | .819** | .836** | **.595**** |

*Note*. * $p < .05$; ** $p < .01$

Correlations of outcome measures within the same rating type (painfulness or unpleasantness) and event type (painful or neutral) are presented on the diagonal and coloured blue. In Experiment 1, all the correlations are > .8 (somewhat higher on average for painful events), indicating that there is a strong relationship between ingroup, rival outgroup, and neutral outgroup ratings within each rating x event combination. In other words, subjects that considered the needle penetrating an ingroup's cheek moderately painful or moderately unpleasant would probably do so for both outgroups; the same stands for q-tips. In Experiment 2 these correlations are also moderate to strong but somewhat lower than in Experiment 1.

Correlations of painfulness and unpleasantness ratings within event type (painful or neutral) are coloured red. Correlations of the same ratings in different event types (painful and neutral separately) are coloured green. Painfulness and unpleasantness ratings within events are moderately positively related: for both ingroups, rival outgroups, and neutral outgroups, events rated as more painful were also rated as more self-unpleasant. However, our participants did have a moderate general tendency to give high or low ratings for pain intensity or self-unpleasantness across event types, as indicated by low to moderate correlations within ratings. In both experiments, correlations within event type are stronger than correlations within ratings; in other words, painfulness and unpleasantness ratings for the same event are more strongly related than painfulness or unpleasantness ratings across events. However, correlations are lower on average in Experiment 2 compared to Experiment 1.

Correlations of different ratings in different event types are coloured purple. These tell us about participants' general tendency to give high or low ratings; they are low and some non-significant in Experiment 2.

As all the dependent variables for neutral events and the majority of variables for painful events were far from normally distributed, we first tried winsorizing the extreme values by using median absolute deviation (MAD), a robust measure of variability, insensitive to both sample size and outliers (Leys, Ley, Klein, Bernard, & Licata, 2012). We replaced all the values 2.5x MAD above or below the median with that value. However, the winsorization left the painful stimuli average ratings virtually unchanged and fixed 7/12 of neutral stimuli ratings to a constant (i.e. the average value of 1 which corresponds to *not at all*) – all the self-unpleasantness ratings were fixed to 1 plus one pain intensity rating (Appendix F).

However, as the ratings for neutral events were very low in general, we approached the data analysis with several different analytical strategies:

First, we analysed the raw (non-winsorized) data, as we had initially planned. Sensitivity analysis for $\alpha = .05$ and desired power .80 indicates that sample size in both Experiments was sufficient to detect small effects for the 3x2 repeated measures interaction (Experimet 1: $\eta^2_p = .032$; Experimet 1: $\eta^2_p = .033$) (MorePower 6.0.4, Campbell & Thompson, 2012).

In each experiment, we analysed the pain intensity and self-unpleasantness ratings separately via two repeated measures 3 x 2 ANOVAs, with the group (ingroup, outgroup, distant outgroup) and stimulus type (painful, neutral) as factors.

In addition, to address the non-normal data distributions, we calculated pain-neutral difference scores for ingroup, rival outgroup and distant outgroup and compared them with two univariate ANOVAs, for pain intensity and self-unpleasantness separately.

Although simulation studies show that RM ANOVA is robust to violations of normality when sphericity assumptions are met (Blanca, Arnau, García-Castro, Alarcón, & Bono, 2023), this was not the case with our data. Although we reported parameters corrected for violations of sphericity, we wanted to further validate or dispute our findings by analysing the winsorizied data and by using the appropriate non-parametric test.

After winsorizing the data, we analysed only ratings for painful events, comparing pain intensity and self-unpleasantness ratings of ingroup, outgroup and distant outgroup models with univariate RM anova.

In addition, we performed a Friedman test, non-parametric test for several related samples for both painful and neutral stimuli separately (as there is no non-parametric match for RM ANOVA with more than one factor).

As all the analyses yielded practically identical results, we report only the full two-way ANOVAs with stimulus type and group as factors on raw data. The remaining analyses are described in Appendix G.

**Experiment 1 – Painful Events.**

For pain intensity ratings, Mauchly's W was significant for both the main effects of group (ingroup, rival outgroup, or neutral outgroup), and the interaction effect ($W_{group} = .765$, $\chi^2(2) = 38.89$, $p = .000$; $W_{interaction} = .790$, $\chi^2(2) = 34.27$, $p = .000$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .810$ for the main effect and $\varepsilon = .826$ for the interaction effect. The main effect of group was not significant, $F(1.62, 236.39) = 1.61$, $p = .206$, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of event type (painful or neutral), the main effect was significant, $F(1, 146) = 86.81$, $p = .000$, $\eta^2_p = .373$, with painful events rated as significantly more painful than neutral on average ($M_p = 2.68$, $M_n = 1.61$). The interaction effect was not significant, $F(1.65, 241.22) = .05$, $p = .929$), indicating that participants' pain intensity ratings of painful and neutral events did not differ with respect to the group identity of the model.

For self-unpleasantness ratings, for the main effect of group (ingroup, rival outgroup, or neutral outgroup), Mauchly's W was significant for both the main effects of group and the interaction effect ($W_{group} = .809$, $\chi^2(2) = 30.76$, $p = .000$; $W_{interaction} = .746$, $\chi^2(2) = 42.42$, $p = .000$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .840$ for the main effect and $\varepsilon = .798$ for the interaction effect. The main effect of group was significant, $F(1.68, 245.14) = 4.55$, $p = .016$, $\eta^2_p = .01$, indicating that participants rated the self-unpleasantness for ingroups, rival outgroups, and neutral outgroups differently ($M_{ig} = 2.03$, $M_{og} = 2.23$, $M_{dog} = 2.09$). Post hoc tests with Bonferroni correction indicated only the difference between ingroups and outgroups was significant (mean difference = -.199, $p = 0.47$), suggesting higher self-unpleasantness ratings for outgroup models compared to ingroup models on average. For the main

effect of event type (painful or neutral), the main effect was significant, $F(1, 146) = 65.19$, $p = .000$, $\eta^2_p = .309$, with painful events rated as significantly more self-unpleasant than neutral on average ($M_{painful} = 2.62$, $M_{neutral} = 1.61$). The interaction effect was not significant, $F(1.60, 232.92) = 2.16$, $p = .128$), indicating that participants' self-unpleasantness ratings for painful and neutral events did not differ with respect to the group identity of the model.

**Experiment 2 – Painful Facial Expressions.**

For pain intensity ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup or neutral outgroup) ($W_{group} = .774$, $\chi^2(2) = 35.30$, $p = .000$;) but not for interaction effect ($W_{interaction} = .991$, $\chi^2(2) = 1.24$, $p = .539$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .816$ for the main effect and $\varepsilon = .991$. Although Mauchley's W for the interaction was not significant, as this was the only case in four separate analyses, we used the Greenhouse-Geisser correction to make all the analyses uniform. The main effect of group was significant, $F(1.63, 226.81) = 3.80$, $p = .032$, $\eta^2_p = .02$, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups differently on average. However, post hoc tests with Bonferroni correction indicated no difference between the means was statistically significant ($M_{ig} = 2.77$, $M_{og} = 2.93$, $M_{dog} = 2.87$). For the main effect of event type (painful or neutral), the main effect was significant, $F(1, 139) = 707.08$, $p = .000$, $\eta^2_p = .836$, with painful events rated as significantly more painful than neutral on average ($M_p = 4.18$, $M_n = 1.53$). The interaction effect was not significant, $F(1.98, 275.54) = .45$, $p = .638$), indicating that participants' pain intensity ratings of painful and neutral events did not differ with respect to the group identity of the model.

For self-unpleasantness ratings, for the main effect of group (ingroup, rival outgroup or neutral outgroup), Mauchly's W was significant for both the main effects of group and the interaction effect ($W_{group} = .585$, $\chi^2(2) = 74.09$, $p = .000$; $W_{interaction} = .946$, $\chi^2(2) = 7.61$, $p = .022$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .706$ for the main effect and $\varepsilon = .949$ for the interaction effect. The main effect of group was significant, $F(1.41, 196.41) = 8.57$, $p = .001$, $\eta^2_p = .06$, indicating that participants rated the self-unpleasantness for ingroups, rival outgroups, and neutral outgroups differently ($M_{ig} = 1.97$, $M_{og} = 2.27$, $M_{dog} = 2.13$). Post hoc tests with Bonferroni correction indicated both the difference between ingroups and outgroups was significant and the difference between ingroups and neutral outgroups was significant (IG-OG mean difference = -.304, $p = .005$, IG-DG mean difference = -.160, $p = .029$), suggesting higher self-unpleasantness ratings for both rival outgroup and neutral outgroup compared to ingroup models on average. For the main effect of event type (painful or neutral), the main effect was significant, $F(1, 139) = 117.21$, $p = .000$, $\eta^2_p = .457$, with painful events rated as significantly more self-unpleasant than neutral on average ($M_{painful} = 2.74$, $M_{neutral} = 1.51$). The interaction effect was not significant, $F(1.89, 263.82) = 1.78$, $p = .173$), indicating that participants' self-unpleasantness ratings of painful and neutral events did not differ with respect to the group identity of the model.

***Intergroup Empathy Bias and Identity, Trait Empathy, and Prejudice***

Descriptive parameters for fan identity measures (Sport spectator identification scale, BIRG-ing, and CORF-ing scales), trait empathy measures – Fantasy, Perspective taking, Empathic concern and Personal distress scales from the IRI, as well as prejudice-related measures – Social dominance orientation and Implicit associations test for the whole sample are presented in Table 15. In addition, we performed independent samples t-tests to see if there was a significant difference

between Red Star and Partizan fans on all measures. The values of t statistic and df as well as the significance of the differences are presented in the last column.

Table 15

*Fan identity, trait empathy, and prejudice – descriptives and Red Star-Partizan fans' differences*

| | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIS | BIRG-ing | CORF-ing | F | PT | EC | PD | SDO | IAT (D) |
| N | 204 | 204 | 204 | 199 | 199 | 199 | 199 | 199 | 191 |
| Min | 4.00 | 1.00 | 1.00 | 1.43 | 1.29 | 1.29 | 1.00 | 1.06 | -1.06 |
| Max | 8.00 | 5.00 | 5.00 | 4.86 | 4.86 | 4.71 | 4.00 | 6.19 | 1.26 |
| Mean | 6.50 | 2.79 | 2.62 | 3.19 | 3.33 | 3.42 | 2.38 | 3.52 | 0.10 |
| SD | 1.24 | 0.77 | 0.89 | 0.68 | 0.59 | 0.59 | 0.66 | 0.94 | 0.55 |
| zSk | -2.79 | 1.89 | 0.80 | 0.07 | -0.27 | -0.36 | 0.14 | -1.88 | -0.85 |
| ZKu | -2.90 | -0.79 | -1.83 | -0.56 | 1.18 | 0.71 | -1.91 | 0.22 | -2.95 |
| SW | 0.920** | 0.983* | 0.981** | 0.990 | 0.987 | 0.984* | 0.986* | 0.982* | 0.967** |
| α | .86 | .66 | .73 | .66 | .64 | .59 | .68 | .85 | / |
| RS-P t(df) | .28 (202) | .10 (202) | -1.13 (202) | 1.21 (197) | 2.24* (197) | 0.67 (197) | 0.27 (197) | -0.36 (197) | -16.39** (189) |

*Note*. SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; IAT – Implicit associations test

As we had targeted fans for our study and as we had used the Sport spectator identification scale were used to exclude low identifiers, it is not unusual that average scores were very high. However, our participants report moderate levels (slightly above the theoretical midpoint) of BaskingInReflectedGlory and CuttingOffReflectedFailure behaviours following their teams' victory or defeat. Interestingly, the strength of identification was related to BIRGing ($r$ = .496, $p$ = .000) but not CORFing behaviours ($r$ = -0.031, $p$ = .656). However, all three fan identity measures were positively related to the Personal distress scale of the IRI (SSIS: $r$ = .226, $p$ = .000, BIRGing: $r$ = .268, $p$ = .000, CORFing: $r$ = .184, $p$ = .009), and two were positively related to SDO (SSIS: $r$ = .147, $p$ = .038, BIRGing: $r$ = .295, $p$ = .000). In addition, BIRGing measure was negatively related to Empathic concern, $r$ = -.151, $p$ = .033, and Perspective taking, $r$ = -.209, $p$ = .003.

Participants rated themselves above the theoretical mean point on all empathy measures except PD. The scores were close to symmetrical on all scales, but the scale reliabilities were not particularly high. Among trait empathy measures, EC, PT, and F were positively related ($r$s .38-.44, $p$s < .000), but none were related to PD ($p$s > .170).

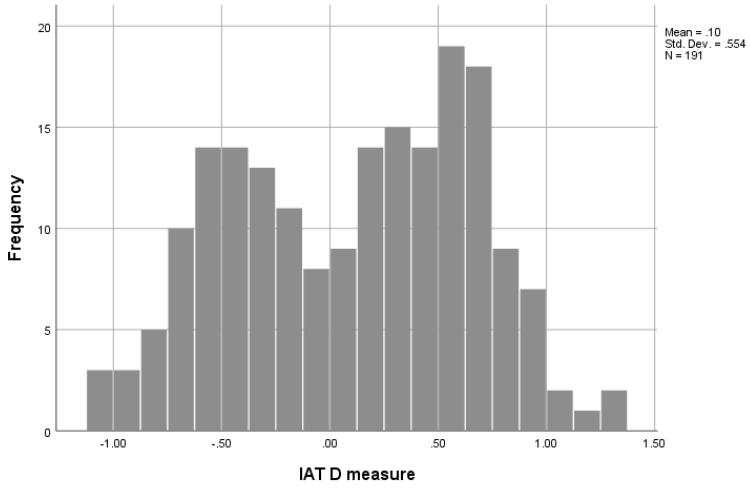
SDO scores were somewhat below the theoretical midpoint and nearly symmetrical, with a high scale reliability. In addition to the positive relationship with fan identity measures, SDO was negatively related to EC, $r$ = -.388, $p$ = .0030 and PT, $r$ = -.315, $p$ = .000, and positively to PD, $r$ = .186, $p$ = .008.

The average IAT D measure was somewhat above 0, which is reasonable because there were more Red Star fans in the sample. The distribution is bimodal, with two groups of scores above and below the mean (Figure 4). Expectedly, Red Star and Partizan fans differ significantly on the D measure. Based on the IAT D measure, we classified the participants into Red Star or Partizan fans and compared that classification to their self-reported preference. Out of 191, 81 participants (42.4%) scored below 0, theoretically showing an implicit preference for Paritzan and 110 participants (57.6%) scored above zero, showing an implicit preference for Red Star. In this sample,

119 participants (62.3%) self-identified as Red Star fans – 84% of them (100 fans) were correctly classified by their IAT score. There were 72 self-identified Partizan fans (37.7%), out of whom 86.1% of them (62 participants) were correctly classified by IAT scores. These results indicate a very good match between the explicit and implicit preferences of the participants. However, the strength of participants' implicit preferences (|D|) did not correlate with any of the fan identity measures, nor with trait empathy and social dominance orientation (Appendix H).

Figure 4

*IAT D measure score distribution for Experiments 1-3*




In summary, our sample consisted of highly identified fans, who can be reliably differentiated based on their IAT scores. However, the strength of their implicit preference towards their own club was neither related to the strength of their explicit identification or fan behaviours, nor their empathic traits and general social attitudes. However, trait personal distress was most reliably related to fan identity measures – fans more prone to personal distress were those who identify more with their team and engage in BIRGing and CORFing behaviours more often. In addition, people higher on SDO in our sample report were more identified with their team and more prone to BIRG.

Spearman's rho of pain intensity and self-unpleasantness ratings in Experiment 1 and Experiment 2 are presented in Table 16 (raw scores) and Table 17 (difference scores painful-neutral).

For raw scores, the trait Personal distress was most reliably correlated with the ratings in general. The correlations were not significant or were low for pain intensity ratings of painful stimuli. However, they tended to be stronger and more reliably present in pain intensity ratings of neutral stimuli, and were all significant and somewhat stronger on average for self-unpleasantness ratings, in all event types and all groups. In absolute values, the correlations of personal distress with ratings were higher in Experiment 2 where participants rated painful and neutral facial expressions. In other words, participants prone to personal distress rated the pain intensity of neutral images somewhat higher and tended to feel more self-unpleasantness both after painful and neutral stimuli. These tendencies were stronger in Experiment 2 on average and were present regardless of the group the target belonged to.

In addition, there were low positive correlations between BIRGing scores and self-unpleasantness ratings of both painful and neutral stimuli for outgroups but not ingroups in Experiment 2.

The remaining correlations are low and sporadic. In Experiment 2, there were some low positive correlations between pain intensity ratings and the remaining three scales of the IRI (F, PT, EC), with more empathic participants providing higher ratings .

Table 16

*Experiments 1 and 2: Correlations of pain intensity and self unpleasantness ratings with fan identity, trait empathy, and prejudice measures*

| rating | group | stimuli | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SSIS | BIRG-ing | CORF-ing | fantasy | pt | ec | pd | SDO | IAT (\|D\|) |
| pain intensity | IG | needle | 0.061 | 0.058 | -0.003 | 0.043 | 0.038 | -0.034 | .191* | -0.068 | -0.080 |
| | OG | needle | 0.060 | 0.023 | 0.008 | 0.012 | 0.046 | -0.015 | 0.149 | -0.126 | -0.045 |
| | DG | needle | 0.029 | 0.049 | -0.011 | 0.017 | 0.012 | -0.012 | .181* | -0.121 | -0.093 |
| | IG | q-tip | 0.053 | 0.113 | 0.002 | -0.051 | -0.037 | 0.060 | .310** | -0.062 | -0.044 |
| | OG | q-tip | 0.101 | 0.058 | -0.077 | -0.065 | -0.003 | 0.058 | .195* | -0.067 | -0.070 |
| | DG | q-tip | 0.089 | 0.114 | -0.019 | -0.110 | -0.015 | 0.015 | .271** | -0.042 | -0.067 |
| self-unpleasantness | IG | needle | 0.082 | 0.084 | 0.101 | 0.042 | 0.019 | -0.024 | .282** | -0.007 | -0.118 |
| | OG | needle | 0.117 | 0.107 | 0.052 | 0.042 | 0.042 | -0.001 | .211* | -0.006 | -0.094 |
| | DG | needle | 0.057 | 0.121 | 0.064 | 0.042 | 0.009 | 0.024 | .261** | -0.022 | -0.083 |
| | IG | q-tip | 0.069 | 0.147 | 0.060 | -0.010 | 0.038 | -0.022 | .294** | -0.026 | -0.081 |
| | OG | q-tip | 0.129 | 0.136 | 0.016 | -0.006 | 0.029 | 0.014 | .219** | -0.003 | -0.083 |
| | DG | q-tip | 0.093 | .165* | 0.092 | -0.094 | 0.019 | 0.014 | .290** | 0.004 | -0.125 |
| pain intensity | IG | painful | 0.112 | 0.087 | -0.141 | .219* | 0.109 | 0.084 | 0.136 | 0.149 | 0.000 |
| | OG | painful | 0.094 | 0.123 | -0.083 | 0.160 | .177* | 0.101 | 0.108 | 0.027 | -0.008 |
| | DG | painful | 0.038 | 0.136 | -0.059 | .199* | 0.118 | 0.055 | 0.128 | 0.144 | -0.089 |
| | IG | neutral | -0.001 | -0.015 | -0.024 | 0.067 | 0.156 | .195* | .225** | -0.035 | -0.008 |
| | OG | neutral | 0.142 | 0.087 | 0.051 | 0.056 | 0.052 | .184* | .292** | 0.067 | -0.006 |
| | DG | neutral | -0.025 | 0.069 | 0.115 | 0.157 | .209* | .189* | .302** | 0.038 | 0.041 |
| self-unpleasantness | IG | painful | 0.072 | 0.156 | 0.138 | 0.063 | 0.011 | 0.102 | .347** | 0.080 | -0.137 |
| | OG | painful | 0.113 | .186* | .189* | 0.063 | 0.008 | 0.085 | .330** | 0.043 | -0.126 |
| | DG | painful | 0.110 | .207* | 0.146 | 0.104 | 0.035 | 0.109 | .322** | -0.007 | -0.121 |
| | IG | neutral | 0.023 | 0.075 | .176* | -0.055 | 0.145 | 0.040 | .317** | -0.017 | -0.045 |
| | OG | neutral | 0.070 | .249** | .189* | -0.063 | 0.034 | 0.108 | .332** | -0.002 | 0.019 |
| | DG | neutral | 0.082 | .277** | 0.152 | -0.060 | 0.049 | 0.078 | .371** | 0.063 | -0.026 |

*Note*. * p < .05; ** p < .01

However, when fan identity measures were correlated to the painful-neutral difference scores, only some correlations with personal distress scores remained significant. Participants more prone to personal distress made a bigger difference in ratings of their own unpleasantness after viewing painful and neutral facial expressions (Experiment 2), for all groups.

Table 17

*Experiments 1 and 2: Correlations of differential pain intensity and self unpleasantness ratings (painful – neutral stimulus for ingroups, rival outgroups and neutral outgroups) with fan identity, trait empathy and prejudice measures*

| rating | group | stimuli | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SSIS | BIRG-ing | CORF-ing | fantasy | pt | ec | pd | SDO | IAT (\|D\|) |
| pain intensity | IG | needle | -0.007 | -0.006 | 0.009 | 0.044 | 0.020 | -0.023 | 0.049 | 0.030 | 0.042 |
| | OG | needle | -0.023 | -0.002 | 0.038 | 0.004 | 0.011 | -0.027 | 0.067 | -0.052 | 0.056 |
| | DG | needle | -0.022 | 0.006 | 0.022 | 0.042 | -0.009 | -0.007 | 0.067 | -0.102 | -0.003 |
| self unpleasantness | IG | needle | -0.009 | -0.047 | 0.055 | 0.024 | 0.005 | 0.009 | 0.092 | 0.019 | -0.050 |
| | OG | needle | -0.074 | -0.082 | 0.027 | 0.040 | 0.071 | -0.026 | 0.096 | -0.060 | -0.119 |
| | DG | needle | -0.048 | 0.009 | 0.010 | 0.090 | 0.044 | 0.043 | 0.135 | -0.050 | -0.063 |
| pain intensity | IG | painful | 0.119 | 0.117 | -0.117 | .184* | 0.004 | -0.016 | 0.031 | 0.157 | -0.011 |
| | OG | painful | 0.003 | 0.035 | -0.132 | 0.130 | 0.124 | -0.022 | -0.096 | -0.006 | -0.025 |
| | DG | painful | 0.014 | 0.072 | -0.093 | 0.127 | 0.025 | -0.102 | -0.054 | 0.095 | -0.039 |
| self unpleasantness | IG | painful | 0.059 | .177* | 0.056 | 0.129 | -0.061 | 0.121 | .235** | 0.081 | -0.145 |
| | OG | painful | 0.025 | 0.052 | 0.059 | 0.109 | -0.003 | 0.026 | .190* | 0.051 | -.187* |
| | DG | painful | 0.078 | 0.116 | 0.117 | 0.165 | 0.041 | 0.046 | .180* | -0.073 | -0.136 |

*Note.* * $p < .05$; ** $p < .01$

In summary, participants more prone to Personal distress tended to give higher ratings in general, but especially for neutral stimuli and for self-unpleasantness ratings. This tendency was stronger in Experiment 2 where the stimuli were directly signalling pain via facial expression. In other words, participants who were more prone to Personal distress also tended to be more personally reactive in general. As for the empathic reaction, i.e. the difference in ratings between painful and neutral stimuli, participants prone to Personal distress were slightly more unpleasant in Experiment 2 after seeing painful expressions of all models, but especially the ingroup.

## Experiment 1 and Experiment 2: Discussion

As discussed in the previous chapter, results of studies of integroup empathy bias for physical pain are not unanimous. While a number of behavioural studies of empathy bias for physical pain report more empathy for the ingroup; the number of studies that do not find ingroup bias is similar. Among neuroscience studies, around a third finds ingroup bias in behavioural empathy ratings, but two-thirds of the studies do not; in contrast, all of them find ingroup bias in neural responses.

In addition to being more numerous in absolute numbers, we argued that neuroscientific findings of differential brain empathic responding to ingroups and outgroups in pain inspired the behavioural intergroup empathy bias studies to a great extent. Therefore, we chose to investigate

whether intergroup empathy can be reliably detected in most frequently used procedures for eliciting empathic response in physical pain studies. When designing the studies, we addressed the typical theoretical and methodological limitations of behavioural measurement in neuroscience we discussed in the previous chapter. We hoped to contribute to the empirical corpus of knowledge about empathy by providing benchmark expectations for these types of empathy-eliciting tasks as well as to provide material for theoretical discussion about behavioural validation of neural indicators of empathic processes. Specifically, we addressed the issues of sample size, physical differences in stimuli, social desirability concerns, and during/post experimental ratings. In addition, by including a third, neutral group, we aimed to provide answers about the mechanism we use to ingroup empathy bias, i.e. whether we increase empathy for "us" or dampen it for "them".

Both of our experiments – the one with painful vs neutral events on neutral faces and the one with painful vs neutral facial expressions – had sufficient power to detect small effect sizes, the samples were significantly larger than the average sample in neuroscience studies, (38 participants, as discussed in the review study) and almost double than the largest sample in any study using similar stimuli (painful facial expressions, 82 participants; Xu, Chen, Kong, & Luo, 2021). We used rival fan groups to operationally define ingroup/outgroup status of the stimuli, with the aim to address two separate issues in current studies that predominantly use racial groups: (a) to eliminate social desirability concerns as much as possible, as fan groups are known to hesitate less in expressing negativity towards their opponents; (b) to eliminate visual differences between stimuli: the stimuli we used to represent ingroups and outgroups were virtually visually identical – the same persons appeared as ingroups, rival outgroups and neutral outgroups to different sets of participants. As we conducted a behavioural study only, participants provided the ratings while viewing the stimuli for the first time, just like in neuroscience studies that collected the explicit ratings "online" i.e. during the scanning procedure. In studies with "online" ratings finding ingroup bias in ratings was somewhat more frequent than in studies that collected ratings offline, after the participants had already been interacting with the stimuli, passively or actively, during the scanning.

However, in both experiments we detected no ingroup bias, neither in pain intensity ratings nor self-unpleasantness ratings. For pain intensity ratings, participants rated the pain of ingroups, rival outgroups and neutral outgroups similarly in both experiments, suggesting the absence of ingroup empathy bias. The results of both experiments also suggest there is no ingroup empathy bias in self-unpleasantness ratings, i.e. the pain-neutral difference is the same for all groups. Participants rated their self-unpleasantness as higher after the rival outgroup's pain compared to the ingroup's (Experiment 1) or both ingroup's and neutral outgroup's (Experiment 2) regardless of the painfulness of the stimuli, i.e. they stated they felt more self-unpleasantness after being presented with a rival outgroup's face regardless of whether it was penetrated by a needle or touched by a q-tip (Experiment 1) or if the face was signaling pain or not (Experiment 2). The effect sizes for this main effect of the group in self-unpleasantness ratings were minuscule and the results cannot be related to any theoretical account of ingroup bias; thus, we find this result is very difficult to interpret meaningfully. In fact, we believe this finding could possibly be an artifact of the simple, decontextualised tasks we presented our participants with. In sum, despite addressing several limitations of previous studies hindering their conceptual and statistical power to detect ingroup empathy bias, our results align with the majority of studies in which ingroup empathy bias in pain intensity and self-unpleasantness was not detected.

Based on the average ratings of pain intensity and self-unpleasantness in both experiments, we argue that the tasks we used failed to elicit a strong reaction in our participants. In Experiment 1 participants rated images of neutral faces with needles signaling painful stimulation and q-tips signaling neutral stimulation – both typical stimuli in this type of research (e.g. Xu et al, 2009; Sheng & Han, 2012). For painful stimuli, both pain intensity and self-unpleasantness ratings were very low – just above 2.5 on a 1-7 Likert scale. For neutral stimuli, the ratings were even lower, just above 1.5 on average. This is a desirable outcome for neutral but not painful stimuli: the observed

difference between painful and non-painful stimuli was very small, indicating that the image of a needle in a model's cheek did not substantially add to the painfulness of the situation (although the differences were statistically significant, an image we rate with 2.5 out of 7 is not a particularly painful nor unpleasant image). This was not the case in previous studies using neutral faces receiving noxious and neutral stimulation. The rating scales in previous studies were variable – from 4 point to 11 point Likert scales. However, the averages seem notably higher than in our study. For example, a study using the same 1-7 Likert scale like we did (Wang et al., 2015) reports painfulness ratings ≈ 5.6/7 for pain intensity and just a bit lower self-unpleasantness ratings (5.1-5.3/7) for painful stimuli. In other studies from the review, the average ratings for painful stimuli were around or above the midpoint of the scale range for both pain intensity and self-unpleasantness.

In contrast to our study, 4 out of 6 previous studies using neutral faces as stimuli presented short 3 second videos instead of still images; it could be that the dynamic nature of the video contributed to the painfulness and self-unpleasantness estimates by making the participants anticipate a painful expression. However, there were no reactions to stimulation (e.g. facial signals of pain or defensive movement) in these videos nor any signs of tissue damage, meaning that the information about painfulness of the stimuli was extracted from the same source in both cases. Moreover, the pain intensity and self-unpleasantness ratings in two studies using still images were also above the midpoint of the scale; therefore, we consider this explanation unlikely.

In Experiment 2 participants rated painful versus neutral facial expressions. Pain intensity ratings in this study were somewhat above the midpoint of the scale for the painful stimuli (≈4.2/7), and very low for the neutral stimuli (≈1.5/7), suggesting participants did recognise the painful facial expressions as such. However, the self-unplesantness ratings were again very low for the painful stimuli – just ≈2.7/7 (and expectedly even lower for the neutral stimuli, ≈1.5/7), suggesting those expressions failed to elicit even a moderate reaction in our participants. Again, this was not the case in previous studies using painful facial expressions, if we look at the average ratings. Although most of them used a 9-point scale, hence making direct comparison with our means impossible, the averages for pain intensity ratings in the vast majority studies were at least a point or two above the midpoint of five (6-7); the averages for self-unplesantness ratings were between ≈4 and ≈5, which is clearly a stronger reaction than elicited in our studies. In the case of Experiment 2, the stimuli in our study and previous studies were structurally identical – still images of painful and neutral facial expressions. However, once again, in our study the participants' reactions to painful stimuli were considerably weaker.

We are far from certain what is the reason for these differences, but we offer several possibilities. First, in previous studies the participants were almost exclusively very young (around 20 years old) university students – arguably more motivated to engage with the task, but traditionally considered to be a WEIRD sample. As the majority of studies was conducted on the East Asian population, the "educated", "industrialised" and "rich" attributes can be applied without discussion. In contrast, although our sample dominantly consisted of students, we had a much broader population reach – our sample included a substantial number of non-students, and students were recruited from faculties teaching study-unrelated subjects. Students, especially those more intimately familiar with general experimental methodology and specific topics under investigation, are expected to approach the experimental situation with a different mindset than an online-recruited volunteer. On one hand, their focus, devotion and the ability and willingness to follow instructions should contribute to the quality of the data; on the other, their answers could be more contaminated with their expectations and previous knowledge. Knowing or being able to make an educated guess about the goals of the experiments as well as participating in an experiment in a laboratory setting (especially a neuroscience one) could motivate participants to behave according to their assigned participant role and steer the ratings towards the upper range. However, this

explanation is tentative and includes accepting several empirically unvalidated expectations as facts; we consider it possible but not particularly certain.

Another reason for weaker reaction to the painful stimuli in our study compared to previous studies might be found in cultural differences. As mentioned, the majority of previous studies (both with painful vs neutral expressions and painful vs neutral events on neutral faces) were conducted on the East Asian population. One of these studies had both Caucasian and Asian samples (Xu et al, 2009), and there was another study comparing Black and Causasian samples in South Africa (Fourie et al, 2017). In both of these studies Caucasian participants had lower average ratings for both ingroup and outgroup targets. The difference is substantial in Xu et al. (2009) study – around 4 points on an 11-point Likert scale for both pain intensity and self-unpleasantness. Studies on emotional face perception point to cross cultural differences in intensity ratings for basic emotions (Ekman et al., 1987, Matsumoto & Ekman, 1989; Matsumoto, 1992); however, compared to East Asian (to be more specific, Japanese) raters, Caucasian raters provided higher, not lower ratings on average for most basic facially displayed emotions, although the relative intensity differences remain. Pain is not traditionally classified as an emotion or studied along basic emotions associated with universally recognised facial expressions. However, pain is readily and reliably detected from facial expressions (Prkachin, 2009; Didline & Atlas, 2019) and there is some preliminary evidence that those expressions are similar within and across cultures (Chen et al., 2016). Although to our knowledge there are no studies comparing pain intensity ratings in others across cultures, based on the emotional face perception literature we believe it is reasonable to hypothesise they might exist. In addition, cultrual influences on various aspects of pain perception, interpretation and reactions to pain have been reliably demonstrated (Peacock & Patel, 2008; Miller & Abu-alhaija, 2019), suggesting that we could also expect them when pain is inferred only from facial expressions of others. Those potential differences could explain why our painful expressions ratings are somewhat lower than in previous studies, as well as to suggest the need to further investigate how cultural differences shape our responses to pain in general (as pain ratings for painful events on neutral faces as well as self-unpleasantness ratigns in both experiments are dramatically lower in our studies).

Taken together, our studies fail to find the hypothesised intergroup bias in pain intensity and self-unpleasantness estimates, neither when those estimates are based on the noxious nature of the event happening on a person's neutral face, nor when they rely upon the painfulness of the facial expression itself. Average ratings indicate that the simple, decontextualised and (in Experiment 1) very artificial stimuli failed to arouse a substantial emotional reaction in most participants. Consequently, as the rating tasks were low-engaging and had little motivational relevance, the situation failed to provoke a strong enough emotional reaction for the effects of motivated social reasoning to arise. Moreover, personal distress was the only trait measure related to pain intensity and self-unpleasantness ratings, but especially to the latter, regardless of the group identity of the model. This suggests that responses were more related to the individual characteristics of the participants than to the social aspects of the task laid before them.

We intentionally avoided to use the phrase "intergroup empathy bias" in our conclusions as we had previously problematised the status of pain intensity estimate as an "empathic reaction". In our opinion, pain intensity ratings (especially when studied in such a decontextualised manner) are conceptually closer to emotion recognition and emotion attributions and could thus be treated as precursors to empathic reactions, but not empathic reactions themselves, as they do not reflect directly the experience of the perceiver aroused by the stimulus but the assessment of the stimulus. Moreover, we would like to point out that, although self-unpleasantness does represent an experience of the perceiver, it is a self-oriented reaction happening as a consequence of a distressing situation, and has the same source but not the same quality as empathic reactions – which are *other-oriented* emotions happening as a consequence of the same situation. We will further discuss the status of these ratings in context of measuring empathy and intergroup empathy bias specifically in general discussion.

In studies that follow, we addressed the artificial decontextualised nature of the situations aimed to arouse empathic reactions by presenting physical pain in circumstances closer to the everyday experience of witnessing other's pain. In addition, we included a broader score of empathic reactions other than pain intensity and self-unpleasantness ratings, to sample the other-oriented spectrum of empathic responses. Specifically, we added empathic concern and perspective-taking as congruent affective and cognitive responses, respectively, as well as an incongruent response (Schadenfreude) hypothesised to represent a counter-empathic response. We investigated if the intergroup empathy bias would emerge in reactions to physically painful situations that could happen to anyone, or if the context needs to be explicitly relevant for the division between *us* and *them.*

## Experiment 3 and Experiment 4: Physical Pain in Context

## Experiment 3: Everyday Misfortunes

In Experiment 3, we presented the participants with short videos of people retelling a painful everyday event that had happened to them. The event had happened either while the person was doing something identity-related (e.g. while going to the football game) or something identity-unrelated (e.g. while going to the store). Participants were asked to rate their empathic reactions to the event. The models were the same people from Experiment 1 and Experiment 2, wearing Red Star, Partizan, or Vojvodina jerseys, representing ingroups, rival outgroups, and neutral outgroups, depending on the participants' identity.

Therefore, in this experiment, we tested whether we could detect intergroup empathy bias for physically painful events by using more complex and contextualised stimuli. Just like in previous experiments:

1. Higher empathy for the ingroup compared to both outgroups would suggest ingroup love as the mechanism of intergroup empathy bias
2. Higher empathy for the ingroup compared only to the rival outgroup would suggest outgroup hate as the mechanism of creating intergroup empathy bias
ratings ordered ingroup – neutral outgroup – rival outgroup would suggest both mechanisms were at stake
3. No differences in ratings between groups would suggest the absence of intergroup empathy bias

Additionally, we expected the effect, if observed, to be stronger in events that happened while the model's action was identity-related, i.e. in events where the models' fan identity was made salient.

### Method

#### *Participants*

Out of 204 participants in this dataset, 121 participants completed Experiment 3 (64 participants after completing Experiment 1, and 57 participants after completing Experiment 2). There were no notable differences in socio-demographic characteristics between this set of participants and the total sample. A more detailed description of the total sample, as well as sample by fan identity and sample by experiment, is presented in Appendix E.

#### *Stimuli and Experimental Tasks*

Twelve different models were presented to participants, with their identities matching the model identities from the previous experiment participants completed. Four models were wearing

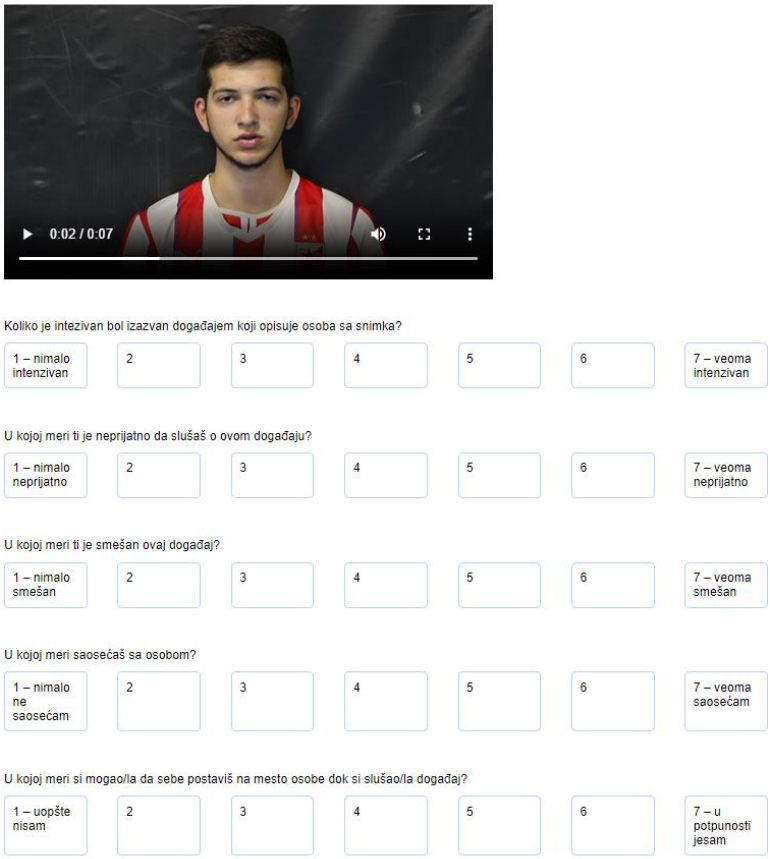
Red Star jerseys, four were wearing Partizan jerseys and four were wearing Vojvodina jerseys, and the ingroup-outgroup status of the stimulus depended on the participants' fan identity, just as in previous experiments.

The task consisted of 12 videos. In each video, a model wearing a jersey retold a painful everyday event that had ostensibly happened to him. In four videos the model was representing the ingroup, in four he was representing the rival outgroup, and in the last four the neutral outgroup. Within each group, in two videos the models were retelling an event where their fan identity was salient, and in the remaining two the same event where their fan identity was not salient, making this a 3(group: ingroup, rival outgroup, neutral outgroup)x2(salience: salient, non-salient) within-subjects study.

Participants were asked to rate the videos for the intensity of the painful event described in the video (pain intensity: *not at all intense – very intense*), how unpleasant they felt while listening about the event (self-unpleasantness: *not at all unpleasant – very unpleasant*), how funny did they find the event (schadenfreude*: not at all funny – very funny*), to what extent did they feel with the person (empathic concern: *not at all – very much*), to what extent did they put themselves in the person's place (perspective taking: *not at all – very much*). Seven-point Likert scale was used for each rating (Figure 5). There was no time restriction for providing ratings. After completing the ratings, they proceeded to the next video.

Figure 5

*Experiment 3: A trial*



Events from the videos were pretested to validate their painfulness and believability; 10 out of 12 events had painfulness ratings M > 5.00 on a 1-7 Likert scale and two were rated below five, and were therefore slightly modified to be perceived as more painful (the full stimuli set with a more detailed description is presented in Appendix C). The salience of the events was manipulated by changing only the keyword or expression pointing to the model's fan identity. To illustrate, the salient version of the event was: *I was walking to buy the tickets **for the game** and I turned my head around because someone called me while continuing to walk and I smashed into the street light.* The key expression was replaced with an identity-neutral one for the non-salient version: *I was walking to buy the tickets **for the gig** and I turned my head around because someone called me while continuing to walk and I smashed into the street light.*

As in previous experiments, subsets of participants saw different model-jersey-event combinations, arranged in Latin square block design to account for potential confounding influences related to the models and the events. There were 6 stimuli sets, depending on the jersey a model was wearing and the salience of his identity in the event. In other words, there were 6 versions of one event for each model – two in each jersey, with one being the salient version and another non-salient version of the event – but each participant saw only one of those.

### *Data Cleaning*

To validate if the participants truly watched the videos and then responded to the ratings, we analysed the response times to the first rating (pain intensity) for each participant. The videos were 6-10 seconds long (M = 7.5 s). As participants needed to click "play" for the video to start, and respond after they had seen it, we took 6000 ms (the length of the shortest videos) as the shortest reasonable time a person needed to rate the pain intensity. We flagged all the trials in which participants had spent less than 6 seconds to provide their first rating and summed them up per participant. In addition, we calculated the average time spent on each video per participant (i.e. the total time spent in watching the video and providing all five ratings).

In total, 40 participants saw 1 or more videos (out of 12) in less than < 6s (out of 12). However, 16 participants provided unexpectedly fast first ratings on 3 or fewer videos, and the remaining 24 participants had 7 and more very fast ratings (8 had 7-9 and 16 participants had 10 or more flagged trials). Moreover, 20 out of these 24 participants were in the bottom 20 on average time spent on video+ratings (the remaining 4 have 9 or 10 flagged videos but spent an unreasonable amount of time on a single one, thereby raising the average time).

As presented in Figure 5, the video and the ratings were presented simultaneously; therefore, we cannot guarantee that the participant didn't preemptively click the rating and then changed it after hearing the whole event; or else, they could provide the self-unpleasantness rating first, and then pain intensity rating, etc. However, our goal was not to clean the data per rating and per trial (as we do not have a clear hypothesis of how much time a person should spend in providing each one) but to identify participants for whom we can believe, with reasonable certainty, that they didn't play the videos at all. Therefore, we decided to exclude the 24 participants who were unreasonably fast in providing their pain intensity ratings more than 50% of the time.

After excluding these participants, the final sample consisted of 97 fans. Sensitivity analysis for $\alpha = .05$ and desired power .80 indicates that this sample was sufficient to detect small to medium effects for the 3x2 repeated measures interaction ($\eta^2_p = .048$) (MorePower 6.0.4, Campbell & Thompson, 2012).

## Results

### *Intergroup Empathy Bias: Main Analyses*

There were 6 ratings per outcome measure (pain intensity, self-unpleasantness, schadenfreude, empathic concern, perspective taking) in each experiment. Descriptive parameters for each rating for salient and non-salient events happening to ingroups, rival outgroups, and neutral outgroups are presented in Table 18.

Pain intensity ratings were slightly above the theoretical midpoint for all scores, indicating the events were perceived as moderately painful. In addition, pain intensity averages were highest of all ratings. Although all scores deviate from the normal distribution according to Shapiro-Wilks test, standardised skewness and kurtosis indicate there were no large horizontal nor vertical distribution asymmetries.

Self-unpleasantness ratings on the other hand were lower on average and positively skewed, indicating that most participants did not consider the events particularly unpleasant to hear about.

However, although the events intentionally resembled internet fail videos to encourage participants to report their feelings, the participants did not find them funny either, as indicated by positively skewed and leptokurtic (for ingroups and rival outgroups) schadenfreude ratings which were lowest of all average ratings.

Empathic concern and perspective taking average ratings are similar and somewhat below the theoretical midpoint, indicating the events provoked moderate to low empathic reactions. Skewness and kurtosis point to no significant assymetries except in scores for rival outgroup empathic concern, which were slightly platykurtic.

Full correlation table (Pearson & Spearman) of all ratings for all outcome measures is presented in Appendix I, and the summary is presented in Table 19. As no measure extremely deviates from the normal distribution, here we discuss Pearson correlations; however, as the number of comparisons is enormous (432 in the full correlation table), we only discuss correlations significant at the .000 threshold.

Correlations of within rating type in different experimental conditions were all significant and positive, moderate to high. Self-unpleasantness, empathic concern and perspective taking ratings had slightly higher intercorrelations compared to pain intensity and schadenfreude.

Within group x salience experimental condition, correlations of pain intensity ratings with both self-unpleasantness ($r$ = .556-.677) and empathic concern ($r$ = .562-.726) were positive and moderate to high. On the other hand, the correlations with perspective taking and schadenfreude were not significant at the corrected threshold.

Self-unpleasantness ratings were moderately positively correlated to empathic concern ($r$ = .396-.547), but not to perspective taking nor schadenfreude ratings. .

Perspective taking and empathic concern were moderately positively related ($r$ = .504-.644 with one correlation not significant at the corrected threshold), but both were unrelated to schadenfreude.

Table 18

*Experiment 3: Empathy ratings for identity salient and non-salient painful events happening to ingroups, rival outgroups and neutral outgroups*

| stimuli | group | R | M | SD | zSk | zKu | SW |
|---|---|---|---|---|---|---|---|
| | | | | pain intensity | | | |
| salient | ingroup | 1-7 | 4.17 | 1.45 | 0.13 | -1.12 | 0.973* |
| | rival outgroup | 1-7 | 4.18 | 1.42 | -0.04 | -0.89 | 0.975 |
| | neutral outgroup | 1.5-7 | 4.14 | 1.27 | 0.64 | -0.85 | 0.973* |
| non-salient | ingroup | 1-7 | 4.35 | 1.45 | -0.23 | -1.64 | 0.973* |
| | rival outgroup | 1-7 | 4.17 | 1.49 | 0.58 | -1.16 | 0.968* |
| | neutral outgroup | 1-7 | 4.21 | 1.37 | 0.72 | -1.32 | 0.971* |
| | | | | self-unpleasantness | | | |
| salient | ingroup | 1-7 | 2.76 | 1.64 | 3.27 | -0.22 | 0.891** |
| | rival outgroup | 1-7 | 2.79 | 1.65 | 3.35 | -0.13 | 0.894** |
| | neutral outgroup | 1-7 | 2.72 | 1.53 | 2.91 | -0.54 | 0.908** |
| non-salient | ingroup | 1-7 | 2.92 | 1.84 | 2.66 | -1.41 | 0.884** |
| | rival outgroup | 1-7 | 2.86 | 1.63 | 2.57 | -0.99 | 0.912** |
| | neutral outgroup | 1-7 | 2.80 | 1.69 | 2.97 | -0.69 | 0.893** |
| | | | | schadenfreude | | | |
| salient | ingroup | 1-7 | 2.10 | 1.34 | 6.34 | 5.01 | 0.798** |
| | rival outgroup | 1-7 | 2.21 | 1.45 | 6.51 | 5.19 | 0.798** |
| | neutral outgroup | 1-5.5 | 2.09 | 1.24 | 4.24 | 0.33 | 0.830** |
| non-salient | ingroup | 1-6 | 2.00 | 1.23 | 5.10 | 2.06 | 0.803** |
| | rival outgroup | 1-7 | 2.19 | 1.46 | 5.41 | 2.43 | 0.799** |
| | neutral outgroup | 1-5.5 | 2.06 | 1.18 | 4.02 | 0.03 | 0.840** |
| | | | | empathic concern | | | |
| salient | ingroup | 1-7 | 3.78 | 1.64 | -0.09 | -1.90 | 0.963** |
| | rival outgroup | 1-7 | 3.44 | 1.60 | 0.41 | -2.11 | 0.951** |
| | neutral outgroup | 1-7 | 3.66 | 1.36 | 0.04 | -0.69 | 0.975 |
| non-salient | ingroup | 1-7 | 3.77 | 1.57 | 0.51 | -1.42 | 0.969* |
| | rival outgroup | 1-7 | 3.62 | 1.57 | 0.16 | -2.03 | 0.962** |
| | neutral outgroup | 1-7 | 3.71 | 1.61 | 0.33 | -1.51 | 0.967* |
| | | | | perspective taking | | | |
| salient | ingroup | 1-7 | 3.87 | 1.69 | -0.01 | -1.71 | 0.959** |
| | rival outgroup | 1-7 | 3.62 | 1.69 | 0.57 | -1.55 | 0.959** |
| | neutral outgroup | 1-7 | 3.76 | 1.65 | 0.36 | -1.33 | 0.966* |
| non-salient | ingroup | 1-7 | 3.85 | 1.77 | 0.91 | -1.93 | 0.951** |
| | rival outgroup | 1-7 | 3.83 | 1.73 | 0.39 | -1.81 | 0.959** |
| | neutral outgroup | 1-7 | 3.65 | 1.66 | 0.23 | -1.56 | 0.959** |

*Note*. M = mean; R – range; SD – standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; SW – Shapiro-Wilks statistic; * $p < .05$, ** $p < .01$

Table 19

*Experiment 3: Pearson correlations between different empathy ratings*

| | PI | SU | SCH | EC | PT |
|---|---|---|---|---|---|
| PI | .430-.660 | | | | |
| SU | .556-.677 | .581-.765 | | | |
| SCH | / | / | .430-.665 | | |
| EC | .562-.726 | .396-.547 | / | .525-.797 | |
| PT | / | / | / | .504-.644 | .585-756 |

*Note*. Correlations of different ratings within experimental condition – off-diagonal, purple; Correlations within empathy ratings in different experimental conditions – on diagonal, green; PI – pain intensity; SU – self-unpleasantness; SCH – Schadenfreude; EC – empathic concern; PT – perspective-taking

To examine whether empathy ratings to painful events differed with respect to the fan group the model belonged to and the salience of fan identity in the story about the event we performed five separate 3(ingroup, rival outgroup, neutral outgroup) x 2(salient, non-salient) repeated measures ANOVAs, with pain intensity, self-unpleasantness, schadenfreude, empathic concern and perspective taking as dependent variables.

For pain intensity ratings, Mauchly's W was not significant for both the main effects of group (ingroup, rival outgroup, or neutral outgroup), and the interaction effect ($W_{group}$ = .989, $\chi^2(2)$ = 1.04, $p$ = .596; $W_{interaction}$ = .945, $\chi^2(2)$ = 5.36, $p$ = .069), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .989 for the main effect and $\varepsilon$ = .948 for the interaction effect. The main effect of group was not significant, $F(1.98, 189.94) = 0.46$, $p$ = .632, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of identity salience (salient or non-salient), the main effect was also not significant, $F(1, 96) = 1.10$, $p$ = .297, pointing to similar average ratings of painful events where the fan identity of the suffering person was salient compared to events where it was not. Lastly, the interaction effect was not significant, $F(1.90, 182.02) = .53$, $p$ = .534), indicating that participants' pain intensity ratings of identity-salient and non-salient events did not differ with respect to the group identity of the model.

For self-unpleasantness ratings, Mauchly's W was not significant for both the main effects of group (ingroup, rival outgroup, or neutral outgroup), and the interaction effect ($W_{group}$ = .966, $\chi^2(2) = 3.33$, $p$ = .189; $W_{interaction}$ = .998, $\chi^2(2) = .16$, $p$ = .921), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .967 for the main effect and $\varepsilon$ = .998 for the interaction effect. The main effect of group was not significant, $F(1.93, 185.61) = 0.29$, $p$ = .741, indicating that participants rated the self-unpleasantness for ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of identity salience (salient or non-salient), the main effect was also not significant, $F(1, 96) = 1.55$, $p$ = .156, pointing to similar average ratings of salient and non-salient events. The interaction effect was not significant, $F(2.00, 191.67) = .78$, $p$ = .882), indicating that participants' self-unpleasantness ratings for identity-salient and non-salient events did not differ with respect to the group identity of the model.

For Schadenfreude ratings, Mauchly's W was significant for the main effects of group (ingroup, rival outgroup, or neutral outgroup), but not for the interaction effect ($W_{group}$ = .899, $\chi^2(2)$ = 10.10, $p$ = .006; $W_{interaction}$ = .999, $\chi^2(2) = .14$, $p$ = .931), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .908 for the main effect and $\varepsilon$ = .999 for the interaction effect. The main effect of group was not significant, $F(1.82, 174.41) = 1.19$, $p$ = .304, indicating that

participants rated the Schadenfreude for ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of identity salience (salient or non-salient), the main effect was also not significant, $F(1, 96) = .70$, $p = .405$, pointing to similar average ratings of salient and non-salient events. The interaction effect was not significant, $F(2.00, 191.71) = .10$, $p = .877$), indicating that participants' Schadenfreude ratings of identity-salient and non-salient events did not differ with respect to the group identity of the model.

For empathic concern ratings, Mauchly's W was significant for the main effects of group (ingroup, rival outgroup, or neutral outgroup), but not the interaction effect ($W_{group} = .917$, $\chi^2(2) = 8.23$, $p = .016$; $W_{interaction} = .994$, $\chi^2(2) = .61$, $p = .735$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .923$ for the main effect and $\varepsilon = .994$ for the interaction effect. The main effect of group was not significant, $F(1.85, 177.29) = 2.94$, $p = .060$, indicating that participants rated the empathic concern for ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of identity salience (salient or non-salient), the main effect was also not significant, $F(1, 96) = .89$, $p = .349$, pointing to similar average ratings of salient and non-salient events. The interaction effect was not significant, $F(1.99, 190.77) = .50$, $p = .604$), indicating that participants' empathic concern ratings for identity-salient and non-salient events did not differ with respect to the group identity of the model.

For perspective taking ratings, Mauchly's W was not significant for both the main effects of group (ingroup, rival outgroup, or neutral outgroup), and the interaction effect ($W_{group} = .947$, $\chi^2(2) = 5.22$, $p = .074$; $W_{interaction} = .984$, $\chi^2(2) = 1.49$, $p = .474$), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon = .949$ for the main effect and $\varepsilon = .985$ for the interaction effect. The main effect of group was not significant, $F(1.90, 182.25) = 1.44$, $p = .240$, indicating that participants rated their perspective taking of ingroups, rival outgroups, and neutral outgroups similarly on average. For the main effect of identity salience (salient or non-salient), the main effect was also not significant, $F(1, 96) = .09$, $p = .767$, pointing to similar average ratings of salient and non-salient events. The interaction effect was not significant, $F(1.97, 189.05) = 1.28$, $p = .280$), indicating that participants' perspective taking ratings of identity-salient and non-salient events did not differ with respect to the group identity of the model.

In summary, neither main effects of group and salience nor their interaction were significant in any of the ratings.

### *Intergroup Empathy Bias and Identity, Trait Empathy, and Prejudice*

Full correlation table of identity-related self report measures, trait empathy measures as well as SDO and IAT D score with ratings in each experimental situation is presented in Appendix J, and the correlations with ingroup, rival outgroup and neutral outgroup ratings averaged for salient and non-salient events are presented in Table 20.

Correlations with fan identity measures, empathic traits and prejudice measures were the strongest and significant at a stricter threshold for empathic concern and perspective taking ratings. However, all the correlations were moderate at best.

For pain intensity ratings, no correlation was significant below .01 threshold, and SDO scores were weakly negatively related to both outgroup ratings at .05.

Table 20

*Correlations of self-report measures with the empathy ratings averaged over salient and neutral events*

| rating | pain intensity | | |
|---|---|---|---|
| group | IG | ROG | NOG |
| SSIS | 0.085 | -0.109 | -0.007 |
| BIRGING | 0.101 | 0.081 | 0.070 |
| CORFING | 0.057 | 0.041 | -0.100 |
| IRI-F | 0.174 | 0.151 | 0.091 |
| IRI-PT | 0.126 | 0.128 | 0.066 |
| IRI-EC | 0.092 | 0.105 | 0.141 |
| IRI-PD | 0.173 | 0.125 | 0.101 |
| SDO | -0.092 | -.238* | -.227* |
| D_ABS | 0.017 | 0.015 | 0.009 |
| rating | self-unpleasantness | | |
| group | IG | ROG | NOG |
| SSIS | 0.197 | 0.050 | 0.141 |
| BIRGING | .210* | 0.122 | 0.174 |
| CORFING | .232* | .213* | .211* |
| IRI-F | 0.085 | 0.025 | -0.035 |
| IRI-PT | 0.007 | 0.099 | -0.045 |
| IRI-EC | 0.041 | 0.073 | 0.082 |
| IRI-PD | .286** | 0.159 | .246* |
| SDO | -0.114 | -0.178 | -.224* |
| D_ABS | -0.007 | -0.041 | -0.028 |
| rating | Schadenfreude | | |
| group | IG | ROG | NOG |
| SSIS | .215* | .261** | .270** |
| BIRGING | .309** | .292** | .328** |
| CORFING | 0.079 | 0.085 | 0.011 |
| IRI-F | -0.150 | -0.037 | -0.028 |
| IRI-PT | -0.062 | -0.017 | 0.002 |
| IRI-EC | -0.148 | -0.081 | -0.033 |
| IRI-PD | 0.041 | 0.179 | 0.139 |
| SDO | 0.171 | .281** | 0.173 |
| D_ABS | 0.008 | 0.197 | 0.069 |
| rating | Empathic concern | | |
| group | IG | ROG | NOG |
| SSIS | 0.009 | -0.153 | 0.068 |
| BIRGING | 0.020 | -0.087 | 0.082 |
| CORFING | 0.076 | 0.048 | 0.063 |
| IRI-F | .231* | 0.198 | 0.112 |
| IRI-PT | .355** | .321** | 0.182 |
| IRI-EC | 0.146 | .295** | 0.204 |
| IRI-PD | 0.037 | -0.031 | 0.105 |
| SDO | -.256* | -.436** | -.327** |
| D_ABS | -0.090 | -0.162 | -0.173 |
| rating | Perspective taking | | |
| group | IG | ROG | NOG |
| SSIS | -.294** | -.347** | -0.192 |
| BIRGING | -0.137 | -0.163 | -0.045 |
| CORFING | 0.163 | .250* | 0.175 |
| IRI-F | .325** | .250* | .246* |
| IRI-PT | .318** | .233* | 0.187 |
| IRI-EC | 0.089 | 0.135 | 0.154 |
| IRI-PD | -0.110 | -0.097 | -0.090 |
| SDO | -.362** | -.423** | -.298** |
| D_ABS | -0.147 | -0.173 | -0.198 |

*Note*. **p < .000; .**p < .01; .*p < .05 ; IG – ingroup, ROG – rival outgroup, NOG – neutral outgroup

For self-unpleasantness ratings, at the most lenient threshold, people more prone to cut off their team after failure tended to give higher ratings for all groups, and people prone to basking in reflected glory of their team after victory only for their ingroup. Personal distress was positively related to ingroup (.01) and neutral outgroup ratings (.05), and SDO was negatively related only to neutral outgroup ratings (.05).

Schadenfreude ratings were positively related to the strength of identification with the preferred team and birging behaviours – interestingly, for both ingroups and outgroups. In addition, higher SDO scores positively predicted schadenfreude for the rival outgroup only.

Participants higher on SDO rate their empathic concern lower for all groups. In contrast, more empathic concern for ingroups and rival outgroups but not neutral outgroups is related to higer scores on perspective taking trait. Fantasy scores are positively related to ingroup empathic concern and empathic concern scores for rival outgroup empathic concern.

Perspective taking ratings have the biggest absolute number of significant correlations with fan identity, trait empathy and prejudice measures. PT ratings were negatively related to SDO for all groups, and to fan identity scores for ingroups and rival outgroups but not neutral outgroups. In contrast, higher fantasy scores were positively related to PT ratings for all groups and higher trait PT for ingroups and rival outgroups, but not neutral outgroups. In addition, corfing scores were positively related to PT ratigns for the outgroups, albeit at the most lenient threshold only.

**Experiment 3: Discussion**

We have previously failed to observe intergroup bias in empathic reactions in tasks most frequently used to investigate intergroup bias in empathy for pain. We argued that the tasks were decontextualised and low-motivating and may have failed to elicit a substantial empathic reaction in participants, not allowing for the effects of group identity to occur or manifest. In this study, we examined whether intergroup bias would emerge in empathic reactions to physically painful events happening in "natural" settings, i.e. if empathic reactions to physical pain presented via complex contextually embedded stimuli would be modified by the identity of the person suffering a small physically painful accident. We presented the participants with video-recorded stories about painful events happening to ingroups, rival outgroups, and neutral outgroups. Additionally, we manipulated the stories by embedding them within fan identity-related versus comparable everyday behaviours, to test whether the salience of the person's fan identity would elicit or modify those reactions.

Our results indicate there was no intergroup bias in any of the empathic reactions assessed: pain intensity, self-unpleasantness, schadenfreude, empathic concern, or perspective-taking. Despite using more ecologically valid stimuli to elicit emphatic reactions and managing to somewhat arouse the participants according to the raw averages, we once again did not observe differences in participants' average ratings of ingroups, rival outgroups, and neutral outgroups. The events were rated as moderately painful; higher-level empathic reactions were rated slightly lower, followed by self-unpleasantness and Schadenfreude.

Absolute values especially for the higher-level empathic reactions tended to be higher for the ingroup ratings (except for schadenfreude in which the rival outgroup condition scores the highest), but the differences were not large nor reliable enough to be statistically significant. Our study had sufficient power to detect small to moderate effect sizes in a 3x2 repeated design.

The results of our studies are inconsistent with previous studies comparable with respect to stimuli and measurement, which warrants further discussion.

Compared to previous studies we reviewed, our study was conceptually similar to the misfortunes studies and to the Trawalter et al (2012) physical pain studies.

a) In misfortunes studies, like in our study, the pain was contextually embedded in a specific event happening to a person who has a certain group identity. Compared to misfortunes studies, in our study the stimuli were focused exclusively on physical pain and presented audiovisually instead of via text (and image). The majority of misfortunes studies measured congruent and incongruent emotional reactions to events ("how good/bad does that make you feel").

b) Trawalter et al. (2012) studies were also focused exclusively on physical pain, like our study, but presented the painful events via single-sentence textual descriptions and asked the participants to imagine those events happening to self, racial ingroups and outgroups. Trawalter studies measured pain intensity of the events.

In addition to pain intensity, we measured several additional empathic reactions, both self and other oriented, congruent and incongruent, covering and expanding this list.

A vast majority of comparable previous studies (all misfortunes studies and one of the two Trawalter studies) registered ingroup empathy bias, eight out of nineteen with a sample size smaller than in this study. Although misfortunes studies were not about physical pain as such, they were unanimous in detecting ingroup bias. Considering we used esimilar empathy-eliciting events (and especially because some stimulus-events in misfortunes studies depict physical pain), this result was reasonable to expect in our study as well, despite we narrowed down the scope of the phenomenon.

If our result was replicated in a high powered study and ones using different groups, it could indicate that physical pain is a too universal experience to be subjected to group biases in empathic reactions at baseline, i.e. when the context of eliciting and measuring empathy does not allow for any direct or indirect identity benefit to the rater brought by biased responding. To elaborate: in our study, participants were instructed to rate the events one by one and the events were not placed in a broader context, i.e. they were not given a cover story explaining the alleged purpose of their ratings. In addition, they reported on their fan identity only after rating the events. Any bias occurring would be interpreted as a spontaneous tendency to empathically react to ingroups and outgroups differently. Our results indicate there is no such spontaneous tendency if we only present physical pain, in any of the empathic reactions measured. However, more research is needed to be able to stand behind this claim.

In Trawalter et al. studies (2012) the painful events were not contextually embedded. In contrast, a portion of misfortunes studies (12/17) were conducted on minimal groups created within a problem solving challenge context. In other words, groups were created within an inherently competitive interaction although being minimal in terms of criteria for ingroup/outgroup divide. This makes it impossible to separate the effects of competition from baseline ingroup empathy bias. In other words, we cannot determine if outgroup membership is sufficient for empathy bias or the competitive interaction with the outgroup is neccessary for the bias to emerge. However, the most of the remaining misfortunes studies investigated differences in baseline empathy (i.e. in an neutral assessment setting but using conflict groups, e.g. national groups in conflict of a variable scale or political groups) similarly to our study, and still found significant ingroup empathy bias.

We do not believe the differences in presentation modality contributed to the differences in findings; on the contrary, we used audiovisual presentation to improve the engagement with the task and directly communicate the group identity via clear visual signals, thereby making the group aspect of the situation visually salient and continuously present. It is plausible to expect these differences to increase the probability to detect intergroup bias, not to reduce it.

Misfortunes studies mostly used unmarked visual analog scales for measuring the congruent (and incongruent) emotional reaction. It could be that this measure is more sensitive/granular and thus more suitable for detecting intergroup bias which could be small and subtle. In our study

participants rated their responses on a 7-point Likert scale, which might have hindered our ability to detect subtle effects. In addition, in misfortunes studies congruent affect was assessed with "how bad does that make you feel?" for negative events. In our study, we labelled the other-oriented emotions rating *empathic concern* and assessed it with "how much do you feel with the person". Despite the terminological confusion about specific terms (e.g. empathic concern and compassion are sometimes used interchangeably and sometimes differentiated), the congruent affect we measured was a higher-level response compared to affective reaction measured in misfortunes studies. It could be that bias reliably exists in spontaneous affect sharing, but disappears as our assessment of the situation becomes more reflective and deliberate.

We placed the physical pain in a realistic everyday context and manipulated the salience of the fan identity of the model suffering pain by manipulating the relevance of the immediate context where the painful event had happened. We detected no ingroup empathy bias in several empathic ratings. However, in this study, although we stressed that the person suffering pain was a fan in one half of the stimuli, the context of assessment was general and disengaged from competitive interactions that form the basis of the rivalry. For example, in an identity-salient event, the model suffered a physical accident while walking to buy the tickets for the football game – it could be any football game, and the outgroup person's suffering could not bring any other benefit to the rater's identity except a vague symbolic one, even if the person missed the game because of it. In the study that follows we investigated ingroup bias in empathic responses to painful events embedded in the context directly relevant to the us/them division, thus making the group membership more salient. Specifically, we investigated empathy for pain of ingroup and outgroup football players suffered during illegal tackles, i.e. fouls.

## Experiment 4: Hey Ref, Foul Play!

In Experiment 4, we presented the participants with images of painful tackles/fouls from real football games and recorded their empathic reactions depending on the identity of perpetrators and victims of the tackle. Just like in previous experiments, the ingroup/outgroup status of the stimuli depended on the participants' identity.

Images of various neutral outgroups were used in this study as well to help dissosicate ingroup love and outgroup hate as the mechanisms of creating bias, if any bias was to occur between ingroups and outgroups' ratings.

### Method

#### *Participants*

Out of 158 participants, 83 (52.5%) identified as Red Star fans and 75 (47.5%) as Partizan fans. Mean age of the participants was ≈ 27, but the median and mode were both below the mean ($M_{age}$ = 27.4 vs $Mdn_{age}$ = 23 vs $Mod_{age}$ = 19). Just like in Experiment 1-3 dataset, the majority of participants were students (61.4%) and rated their socioeconomic status as slightly above average (M = 57.76 on a 1-101 unmarked slider). However, the number of students was smaller than in

previous dataset and a quarter of the sample was older than 35 ($Q_{75}$ = 35), pointing to a slightly more diverse sample. Partizan and Red Star fans did not differ substantially in the ratio of students (Red Star 62.7% vs Partizan 60.0%), and were of similar age, with Partizan fans having a higher percentage of older participants (Red Star $M_{age}$ = 27.0, $Q_{75}$ = 33 vs Partizan $M_{age}$ = 27.9, $Q_{75}$ = 37).

### *Stimuli and Experimental Tasks*

In Experiment 4, participants were presented with photographs of physical clashes taken during real football matches in the Serbian SuperLiga (i.e. the country's primary football competition). The group identity of the fouled player and the player who committed the foul (the perpetrator) was systematically varied: both could belong to the ingroup, to the rival outgroup, or to the neutral outgroup, which resulted in six possible combinations (i.e. experimental situations), shown in Figure 3. As in previous experiments, the ingroup/rival outgroup status of the players varied according to the club affiliation of the participant. For example, a photograph of a Red Star player tackling a Partizan player belongs to rival outgroup pain situation for a Red Star fan, but it is classified as ingroup pain for a Partizan fan. Players of various other Superliga clubs represented the neutral outgroup for both Partizan and Red Star fans.

In each experimental situation, we used 10 photographs to measure the participant's response to different groups, which resulted in 60 stimuli in total. The order of the stimuli was randomised. As the number of stimuli and associated questions was large, the participants were offered to take a break after completing the first half of the task.

Participants were asked to reply 4 different questions about the event depicted in the photograph:

1) control question: the identity of the player in pain (by position on the photograph – left or right, or by jersey colour in photographs where the position of the players couldn't be unambiguously stated). As the stimuli were real photographs, i.e. not designed for experimental purposes, and they were depicting complex and potentially ambigous events (especially if we have the stereotype that football players tend to fake being injured in mind), we included the control question to validate that the participants interpreted the situation the way it was classified in experimental design).

2) pain intensity ("How painful was this tackle for the fouled player?", 7-point Likert scale, from *not at all* to *very painful*)

3) congruent emotional response (empathic concern) ("How sorry you are for the fouled player?", 7-point Likert scale, from *not sorry at all* to *very sorry*)

4) incongruent emotional response/schadenfreude („How joyful does the fouled player's pain make you", 7-point Likert scale, from *not joyful at all* to *very joyful*)

The stimulus exposure time was unlimited, i.e. the photograph was presented on the screen the whole time while participants were rating the event. The ratings were presented sequentially – after providing the first one, the next one would appear, etc.

The photographs were mainly drawn from a professional photo database which is the main photo source for Serbian sports journalists (https://starsport.photoshelter.com/). One dimension of the photo (width or height, depending on the orientation) was fixed to 1200, while the other varied between 750 and 950 pixels. All photographs were high-quality colour photographs. In each photograph, the event of interest dominated the scene and target actors were the primary focus of the photograph (if present, other players were blurred and blended in the background of the scene).

Photographs of football players of several clubs (other than Red Star and Partizan, who represent ingroups and rival outgroups to their respective fan base) were used to represent the neutral outgroup experimental condition because the number of available photographs that fulfill the (very lenient) conditions listed above was insufficient for any specific Superliga club other than Red Star and Partizan. However, this fact alone indicates the dominance of the "eternal rivals" in Serbian football and supports the premise that they are the most important (and only relevant) rival outgroup to each other, and consequently their fan base.

**Results**

Before calculating the mean scores, we first analysed how accuracte the participants were in determining the victim of the foul, i.e. which stimuli were interpreted unambigously. As we presented complex events, participants had to correctly identify which player was suffering pain in each image before providing empathic ratings. Each perpetrator – fouled player combination was represented with 10 stimuli. We counted the stimulus as eligible for inclusion in the total score if more than 90% of the participants had correctly identified the victim of the tackle. Mean accuracy for each combination as well as the number of eligible stimuli (i.e. the number of stimuli where the fouled player was correctly identified in > 90% of the cases) is presented in Table 21. Accuracy for each individual stimulus is presented in Appendix K.

Table 21

*Accuracy analysis*

| perpetrator | fouled player/victim | mean accuracy (10 stimuli) | N of unambiguous stimuli (accuracy > 90%) | mean painfulness – all (accurate responses) |
|---|---|---|---|---|
| Partizan | Red Star | 94.9% | 10 | 3.30 |
| other club | Red Star | 92.4% | 8 | 3.32 |
| Red Star | Partizan | 89.1% | 7 | 3.47 |
| other club | Partizan | 95.3% | 9 | 3.68 |
| Red Star | other club | 81.8% | 5 | 3.56 |
| Partizan | other club | 93.1% | 9 | 3.13 |

*Note*. N = 158

The number of stimuli with high accuracy varied from 50% to 100% per experimental condition, and was somewhat lower when Red Star players were the perpetrators of the foul. In other words, the number of unambiguous images varied substantially per experimental condition. In addition, there were differences between average painfulness ratings between conditions. As the number of Red Star and Partizan fans was approximately equal, we removed the ratings for inaccurate responses and compared the average painfulness of images in each combination regardless of participants' fan identity to check if we could treat these experimental conditions as equal. One-way repeated measures ANOVA ($W = .112$, $\chi^2(2) = 340.05$, $p = .000$, $\varepsilon = .446$) indicated there was a significant difference between conditions ($F(2.23, 350.03) = 22.43$, $p = .000$, $\eta^2 = .125$). Post hoc pairwise comparisons (Bonferroni corrected) indicated that images where other clubs' players tackle Partizan players were seen as significantly more painful than all other conditions, and the images where Partizan players tackle other clubs' players as significantly less painful than all other conditions; in addition, both conditions in which Red Star players were the perpetrators were

seen as significantly more painful than the two conditions where Red Star players were the victims (Appendix L).

These differences prevented us from merging the ratings from two separate conditions into a common variable depending on participants identity, e.g. combining Red Star(P)-Partizan(V) condition for Partizan fans and Partizan(P)-Red Star(V) condition for Red Star fans into a common variable "ingroup pain", like we did in previous experiments. Therefore, we first needed to choose a number of images <10 from each experimental condition that are approximately equal on painfulness. We took only non-ambigous images into consideration (accuracy > 90%), and in conditions where the number of unambiguous images was greater than needed, we aimed to choose the images to be as similar as possible on average painfulness to the images from the limiting experimental condition (i.e. the one with the lowest number of unambiguous images).

In addition, although all participants were presented the same stimuli, repeated measures analysis with perpetrator and victim identity as factors could not be performed as the fully crossed design was not possible – the situations in which both the perpetrator and the victim belong to the same team (i.e. same factor level) are impossible in a real game and conceptually meaningless. We thus employed two somewhat unorthodox analyses:

First, we compared the experimental conditions where ingroups and rival outgroups were the victims of the tackle i.e. they were receiving painful stimulation, depending on the context of the tackle - whether it happened during their direct confrontation i.e. during the derby, or the perpetrator was a member of the third team, i.e. during an "ordinary" game. These conditions were the ones where the rivals alternate as perpetrators and victims during the derby with the rival (Partizan P – Red Star V vs Red Star P – Partizan V) and during the "ordinary" game in which a neutral outgroup player was the perpetrator, and Red Star or Partizan players were the victims. (Other club P – Partizan V vs Other club P – Red Star V). For the conditions where the Red Star and Partizan players were the victims of the tackle, the lowest number of unambiguous images per condition was 7 in the Red Star (P) – Partizan (V) condition with average painfulness 3.55. Therefore, we excluded 3 unambiguous images from Partizan-Red Star condition (with the lowest painfulness rating, the remaining 7 having painfulness rating 3.52), 1 from other club – Red Star (with the lowest painfulness rating, the remaining 7 having average rating 3.46) and 2 from other club – Partizan condition (with the highest painfulness rating, the remaining 7 having the average rating of 3.52).

Second, we compared the experimental conditions where the ingroups and outgroups were the perpetrators of the tackle, and the victims were the members of a third team. For this analysis the lower number of unambiguous images was 5 the Red Star-other club condition with mean painfulness 3.64; therefore, we excluded 4 unambiguous images from the Partizan – other club condition (with the lowest painfulness rating, the remaining 5 having the average rating 3.54).

The descriptive parameters per experimental condition for the final set of stimuli are presented in Table 22. Repeated measures ANOVAs ($W_1$ = .112, $\chi_1^2(2)$ = 340.05, $p$ = .000, $\varepsilon_1$ = .446) comparing the condition regardless of participants' club affiliation indicated there were no significant differences on average painfulness between conditions in both sets of variables ($F_1(1.68, 263.04)$ = .53, $p$ = .557); $F_2(1, 157)$ = 1.18, $p$ = .279), thereby justifying the merging of variables according to ingroup-outgroup status of the victims and perpetrators of the tackle.

Table 22

*Painfulness per experimental condition in the final stimuli set*

| perpetrator | fouled player/victim | N of stimuli | Min | Max | M | SD | zSk | zKu | SW |
|---|---|---|---|---|---|---|---|---|---|
| Partizan | Red Star | 7 | 1.00 | 6.75 | 3.52 | 1.13 | 1.777 | 0.172 | 0.988 |
| other club | Red Star | 7 | 1.00 | 7.00 | 3.46 | 1.04 | 3.087 | 0.936 | 0.974** |
| Red Star | Partizan | 7 | 1.57 | 7.00 | 3.55 | 1.05 | 4.185 | 1.383 | 0.954** |
| other club | Partizan | 7 | 1.43 | 7.00 | 3.52 | 1.09 | 4.438 | 2.299 | 0.954** |
| Red Star | other club | 5 | 1.75 | 7.00 | 3.65 | 1.14 | 3.398 | 0.376 | 0.957** |
| Partizan | other club | 5 | 1.00 | 6.25 | 3.55 | 1.02 | 0.518 | -0.038 | 0.992 |

*Note*. N = 158; Min – minimum; Max – maximum; M – mean; R – range; SD – standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; SW – Shapiro-Wilks statistic; * p < .05, ** p < .01

### *Rivals as Victims of Foul Play*

Descriptive parameters for pain intensity, comapassion and schadenfreude ratings for ingroups and rival outgroups depending on the perpetrator are presented in Table 23.

Table 23

*Painfulness, empathic concern and schadenfreude for ingroups and rival outgroups depending on the perpetrator*

| rating | perpetrator | victim | Min | Max | M | SD | zSk | zKu | SW |
|---|---|---|---|---|---|---|---|---|---|
| painfulness | rival outgroup | ingroup | 1.60 | 7.00 | 3.79 | 1.06 | 2.302 | -0.405 | 0.979** |
| | other | ingroup | 1.43 | 7.00 | 3.69 | 1.09 | 3.605 | 0.944 | 0.965** |
| | ingroup | rival outgroup | 1.00 | 6.57 | 3.27 | 1.05 | 3.828 | 2.925 | 0.959** |
| | other | rival outgroup | 1.00 | 7.00 | 3.29 | 1.00 | 4.027 | 2.788 | 0.964 |
| empathic concern | rival outgroup | ingroup | 1.00 | 7.00 | 3.65 | 1.70 | 1.516 | -2.357 | 0.958** |
| | other | ingroup | 1.00 | 7.00 | 3.51 | 1.68 | 2.357 | -1.900 | 0.952** |
| | ingroup | rival outgroup | 1.00 | 7.00 | 2.53 | 1.40 | 5.252 | 2.130 | 0.900** |
| | other | rival outgroup | 1.00 | 7.00 | 2.67 | 1.46 | 5.012 | 1.618 | 0.907** |
| schadenfreude | rival outgroup | ingroup | 1.00 | 5.50 | 1.17 | 0.56 | 24.714 | 71.653 | 0.346** |
| | other | ingroup | 1.00 | 7.00 | 1.19 | 0.65 | 31.361 | 119.212 | 0.320** |
| | ingroup | rival outgroup | 1.00 | 7.00 | 1.91 | 1.71 | 10.192 | 6.895 | 0.595** |
| | other | rival outgroup | 1.00 | 7.00 | 1.69 | 1.39 | 12.850 | 15.050 | 0.568** |

*Note*. N = 158; Min – minimum; Max – maximum; M – mean; R – range; SD – standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; SW – Shapiro-Wilks statistic; * p < .05, ** p < .01

Firstly, we compared fans' rating of physical pain intensity in the images where the rivals alternate as victims (victim: ingroup vs outgroup) depending on the perpetrator of the foul – a member of neutral outgroup versus their own ingroup/rival outgroup. The main effect of victim identity was significant, $F(1, 157) = 47.62$, $p = .000$, $\eta^2_p = .233$) with tackles where ingroups were the victims rated as more painful than those where rival outgroups were the victims. However,

neither the identity of the perpetrator, $F(1, 157) = 1.92$, $p = .168$, nor their interaction, $F(1, 157) = 3.58$, $p = .060$, were significant.

Secondly, we compared fans' empathic concern ratings for the ingroup and rival outgroup victim in question depending on the perpetrator. The main effect of victim identity was significant, $F(1, 157) = 65.88$, $p = .000$, $\eta^2_p = .296$) with participants reporting significantly more empathic concern for ingroups compared to outgroups. The identity of the perpetrator was not significant, $F(1, 157) = .01$, $p = .929$, with derby games (ingroup and rival outgroup as perpetrators/victims) and other games (neutral outgroup as perpetrators and the rival teams as victims) rated similarly on average. However, the interaction of the victim and perpetrator was significant, $F(1, 157) = 19.01$ $p = .000$, $\eta^2_p = .108$. Post hoc comparison with separate paired samples t-tests indicated that both ingroup ($t(157) = 3.16$, $p = 000$) and rival outgroup ($t(157) = -3.10$, $p = 000$) ratings differ significantly with respect to the perpetrator, but the direction of the difference is reversed: empathic concern for the ingroup was greater if they were fouled by the rival outgroup (in the derby) than by the neutral outgroup, but empathic concern for the outgroup was greater if they were fouled by the neutral outgroup compared to their own ingroup.

Figure 6

*Pain intensity ratings for ingroup and outgroup victims depending on the perpetrator*

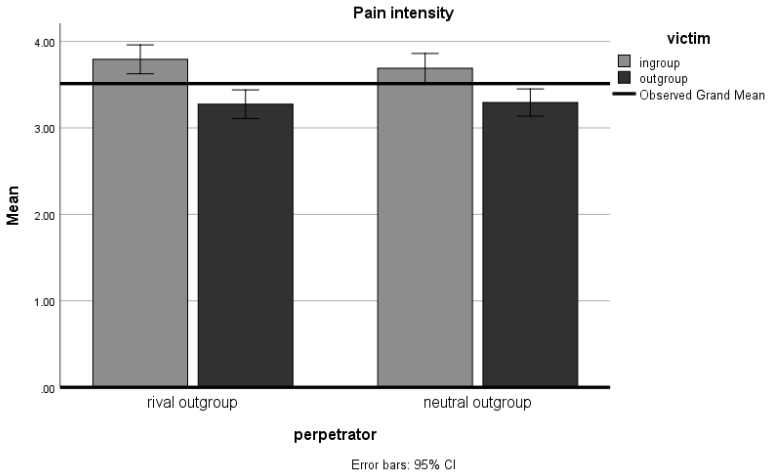
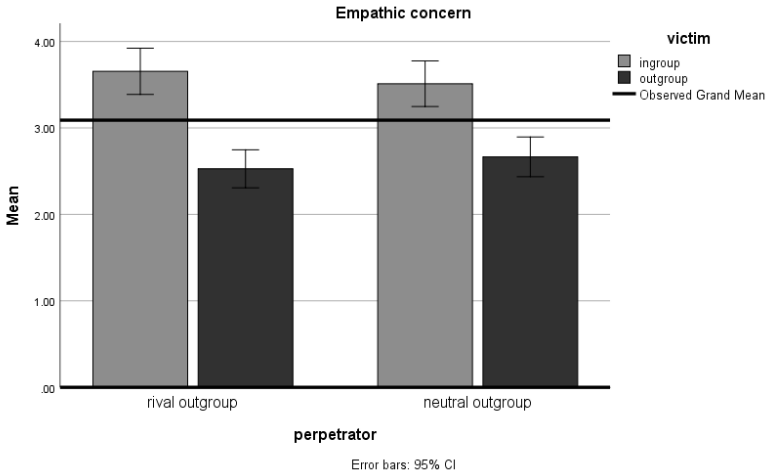



Figure 7

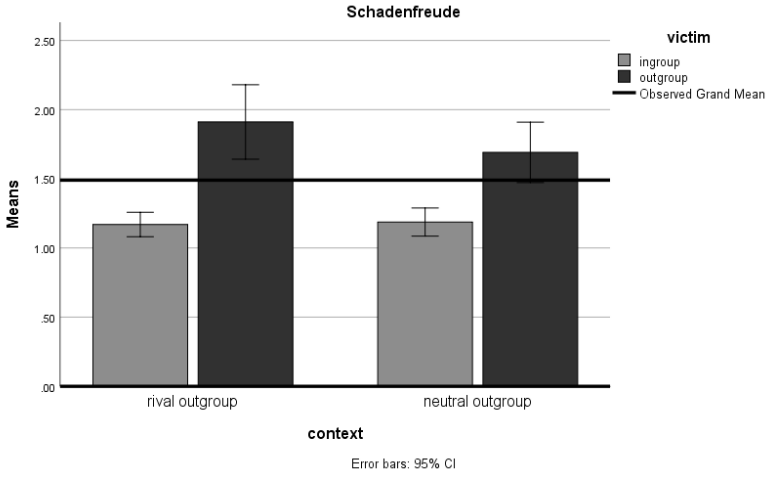
*Empathic concern ratings for ingroup and outgroup victims depending on the perpetrator*



Lastly, we compared fans' schadenfreude for the victim depending on both his identity and the identity of the perpetrator. The main effect of victim identity was significant, $F(1, 157) = 32.92$, $p = .000$, $\eta^2_p = .173$) with participants feeling more happy when seeing tackles with the rival outgroup as the victim compared to ingroup, on average. The identity of the perpetrator was also significant, $F(1, 157) = 13.84$, $p = .000$, $\eta^2_p = .081$ with derby takcles rated higher than tackles by neutral outgroups on average. In addition, the interaction of the victim and perpetrator was significant, $F(1, 157) = 18.22$, $p = .000$, $\eta^2_p = .104$. Post hoc comparison with separate paired samples t-tests indicated that only rival outgroup ($t(157) = 4.28$, $p = 000$) but not ingroup ($t(157) = -.89$, $p = .374$) ratings differ significantly with respect to context, with participants reporting significantly more schadenfreude for the outgroup in the context of the derby, i.e. when their own ingroup was the perpetrator of the tackle.

Figure 8

*Scahdenfreude ratings for ingroup and outgroup victims depending on the perpetrator*




***Rivals as Perpetrators of Foul Play***

Descriptive parameters for pain intensity, comapassion and schadenfreude ratings for the neutral outgroup depending on the ingroup versus rival outgroup perpetrator are presented in Table 24.

We compared fans' rating of physical pain intensity of neutral outgroup victims' pain in the images where the rivals alternate as perpetrators with one-way repeated measures ANOVA. The main effect of perpetrator identity was significant, $F(1, 157) = 9.81$, $p = .000$, $\eta^2_p = .059$) with tackles where ingroups were the perpetrators rated as less painful than those where rival outgroups were the perpetrators.

Secondly, we compared fans' empathic concern ratings for the neutral outgroup victim depending on the perpetrator. The main effect of perpetrator identity was significant, $F(1, 157) = 19.36$, $p = .000$, $\eta^2_p = .110$) with participants reporting significantly more empathic concern for the victim when the rival outgroup had hurt them compared to the ingroup.

Lastly, we compared fans' schadenfreude for the victim depending on the perpetrator. The main effect of victim identity was significant, $F(1, 157) = 13.46$, $p = .000$, $\eta^2_p = .079$) with

participants feeling more happy when seeing tackles with the ingroup as the perpetrator compared to the rival outgroup.

Table 24

*Painfulness,* empathic concern *and schadenfreude for the neutral outgroup depending on the perpetrator (ingroup or rival outgroup)*

| rating | perpetrator | Min | Max | M | SD | zSk | zKu | SW |
|---|---|---|---|---|---|---|---|---|
| pain intensity | ingroup | 1.00 | 7.00 | 3.46 | 1.08 | 1.501 | -0.303 | 0.988 |
| | rival outgroup | 1.60 | 7.00 | 3.74 | 1.07 | 3.234 | 0.864 | 0.969** |
| empathic concern | ingroup | 1.00 | 7.00 | 2.98 | 1.49 | 3.694 | -0.361 | 0.939** |
| | rival outgroup | 1.00 | 7.00 | 3.38 | 1.50 | 2.766 | -0.933 | 0.962** |
| echadenfreude | ingroup | 1.00 | 7.00 | 1.41 | 0.96 | 18.609 | 40.035 | 0.490** |
| | rival outgroup | 1.00 | 6.00 | 1.21 | 0.66 | 23.627 | 63.419 | 0.369** |

*Note*. N = 158; Min – minimum; Max – maximum; M – mean; R – range; SD – standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; SW – Shapiro-Wilks statistic; * p < .05, ** p < .01

***Intergroup Empathy Bias and Identity, Trait Empathy, and Prejudice***

Descriptive parameters for fan identity measures (Sport spectator identification scale, BIRG-ing, and CORF-ing scales), trait empathy measures – Fantasy, Perspective taking, Empathic concern and Personal distress scales from the IRI, as well as prejudice-related measures – Social dominance orientation and Implicit associations test for the whole sample are presented in Table 25. Like in previous dataset, we compared the scores of Red Star and Partizan fans on all measures (t statistic, dfs and significance presented in the last row).

Similarly to previous dataset, the scores on Sport spectator identification scale were very high, but BIRGing and CORFing frequency moderate, with SSIS being related to BIRGing ($r$ = .550, $p$ = .000) but not CORFing ($r$ = -0.013, $p$ = .876). SSIS and BIRGing were also related to SDO (SSIS: $r$ = .223, $p$ = .000, BIRGing: $r$ = .258, $p$ = .000), like in previous sample. BIRGing was positively related to Fantasy ($r$ = .179, $p$ = .025) and CORFing to personal distress ($r$ = .200, $p$ = .012), but both correlations were very low and significant only by most lenient criteria.

Trait empathy measures behaved similarly as in the previous sample, with mean socres above the theoretical mean point on all measures except PD, symmetrical distributions and not particularly high reliabilities. Also, the pattern of relationships between the mesures was similar, with EC, PT, and F being positively related (*r*s .32-.46, *p*s < .000), but none being related to PD (*p*s > .06).

Similarities with the previous sample were evident with respect to SDO as well, with symmetrical scores somewhat below the theoretical midpoint and high scale reliability. Except for positive relationships with fan identity measures, SDO was negatively related to EC, $r$ = -.255, $p$ = .000 and PT, $r$ = -.465, $p$ = .000, just like in previous sample. In this sample, however, the relationship to F was significant, $r$ = -.263, $p$ = .001, but the relationship to PD was not significant this time, $r$ = .026, $p$ = .750.

Table 25

*Fan identity, trait empathy, and prejudice – descriptives and Red Star-Partizan fans' differences*

| | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIS | BIRG-ing | CORF-ing | F | PT | EC | PD | SDO | IAT (D) |
| N | 158 | 158 | 158 | 158 | 158 | 158 | 158 | 158 | 140 |
| Min | 3.29 | 1.33 | 1.00 | 1.57 | 1.57 | 1.86 | 1.00 | 1.00 | -2.00 |
| Max | 8.00 | 5.00 | 5.00 | 4.86 | 4.71 | 5.00 | 3.86 | 5.31 | 1.16 |
| Mean | 5.91 | 2.77 | 2.43 | 3.30 | 3.41 | 3.47 | 2.41 | 3.07 | -0.03 |
| SD | 1.18 | 0.78 | 0.86 | 0.66 | 0.60 | 0.61 | 0.64 | 1.03 | 0.63 |
| zSk | -1.023 | 1.764 | 2.468 | -0.348 | -1.189 | -0.519 | -0.778 | -0.467 | -1.624 |
| ZKu | -2.312 | -0.965 | -0.305 | -0.884 | -0.033 | 0.056 | -1.152 | -1.895 | -1.010 |
| SW | 0.971** | 0.979* | 0.969** | 0.987 | 0.989 | 0.991 | 0.983 | 0.984 | 0.956** |
| α | .79 | .70 | .74 | .70 | .67 | .61 | .71 | .86 | / |
| RS-P t(df) | -2.44* (156) | -.66 (156) | -.36 (156) | .77 (156) | .42 (156) | -.06 (156) | -.78 (156) | -1.58 (156) | 15.84** (138) |

*Note*. SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; IAT – Implicit associations test; Min – minimum; Max – maximum; M – mean; R – range; SD – standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; SW – Shapiro-Wilks statistic; * $p < .05$, ** $p < .01$

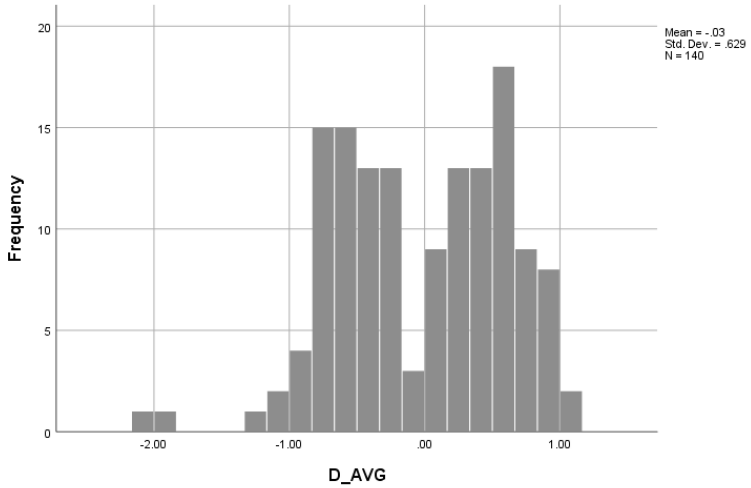
The average IAT D measure was close to 0 and the distribution bimodal (Figure 9), which was expected because the number of Red Star and Partizan fans in the sample was similar. Also, as expected, fans differed significantly on their implicit preferences. We again classified the participants into Red Star of Partizan fans based on their D score and compared the classification to their self-report identity. Out of 140 participants, 68 participants (48.6%) scored below 0, theoretically showing an implicit preference for Paritzan and 72 participants (51.4%) scored above zero, showing an implicit preference for Red Star. In this sample of 140 participants with IAT scores, 76 participants (54.3%) self-identified as Red Star fans – 88.2% of them (67 fans) were correctly classified by their IAT score. There were 64 self-identified Partizan fans (45.7%), out of whom 92.2% of them (59 participants) were correctly classified by IAT scores. Classification accuracies once again point to a very good match between the explicit and implicit preferences of the participants. In addition, in this sample the strength of participants' implicit preferences (|D|) correlated with their SSIS socre ($r = .338$, $p = .000$) (and with F score, but this correlation is very low and borderline significant, $r = .172$, $p = .042$).

This sample also consisted of highly identified fans, who can be reliably differentiated based on their IAT scores, but whose implicit and implicit preferences this time match. Just as in previous sample, SDO was related to both explicit identification and BIRGing. However, the pattern of relationships with trait empathy measures was not repeated.

As most of the ratings deviate from the normal distribution, and in particular because schadenfreude ratings are heavily skewed, we presented the Spearman Rho correlations between trait measures and ratings (both raw and difference scores) in Table 26.

Figure 9

*IAT D measure score distribution for Experiments 1-3*




**Rivals as Victims of Foul Play.** Among fan identity measures, team identification and BIRGing were related to both raw and difference scores, but CORFing was not. For raw scores, participants who were more highly identified with their team and who were more prone to BIRG rated the pain of their ingroup as more intense, had more empathic concern for the ingroup and less for the rival outgroup, and had higher scores on schadenfreude for their rival outgroups' pain, for both derby and non-derby tackles (although the correlations tended to be stronger and significant at a stricter threshold for the derby tackles). Moreover, the difference between ingroup and rival outgroup pain ratings, empathic concern and schadenfreude in both derby and non-derby tackles was more strongly related to fan identity and BIRGing behaviour.

In contrast, the correlations of both raw and difference ratings with trait empathy were low, negative, non-systematic and and the vast majority was significant only at the most lenient threshold. Participants with higher scores on personal distress tended to provide higher ratings for the outgroups' tackles, and participants higher on perspective taking reported less shadenfreude in some situations.

Social dominance orientation was positively related to 3/4 raw shadenfreude ratings as well as 5/6 ingroup-outgroup differences (in pain intensity, empathic concern and schadenfreude). These effects were, however, rather low.

Participants' implicit preference towards their own team was positively related to raw ingroup empathic concern regardless of the context of the tackle (derby or non-derby), as well as to all ingroup-outgroup differences in pain intensity, empathic concern and schadenfreude, in both contexts.

**Rivals as Perpetrators of Foul Play**. When their team member was inflicting pain in a tackle to a neutral outgroup, neither raw pain intensity and empathic concern ratings nor difference scores were related to any of the fan identity, trait empathy or prejudice measures. Raw and difference schadenfreude ratings were related to some of these measures but the correlations were low, sporradic and significant only at the most lenient threshold (with two exceptions: (1) a positive relationship between SDO and raw schadenfreude for the neutral outgroup when the ingroup was the perpetrator as well as with the difference score; (2) a negative relationship of the schadenfreude difference between ingroup and rival outgroup perpetrator and personal distress – participants more prone to personal distress made a smaller difference in schadenfreude for the neutral outgroup depending on the perpetrator).

Table 26

*Correlations between pain intensity, empathic concern and schadenfreude ratings with fan identity, trait empathy, and prejudice*

| rating | perpetrator | victim | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SSIS | BIRG-ing | CORF-ing | F | PT | EC | PD | SDO | IAT (\|D\|) |
| pain intensity | ROG | IG | .237** | .265** | 0.137 | 0.084 | -0.024 | -0.084 | 0.034 | .159* | 0.133 |
| | IG | ROG | -0.156 | -0.007 | 0.082 | 0.095 | -0.041 | -0.086 | 0.106 | 0.019 | -0.029 |
| | NOG | IG | .191* | .244** | 0.105 | 0.154 | 0.062 | -0.049 | 0.017 | 0.149 | 0.125 |
| | NOG | ROG | -0.071 | 0.044 | 0.071 | 0.063 | -0.122 | -0.089 | 0.092 | 0.033 | -0.026 |
| | IG | NOG | -0.085 | 0.056 | 0.022 | 0.047 | -0.035 | -0.103 | 0.099 | 0.014 | -0.025 |
| | ROG | NOG | 0.069 | .172* | 0.106 | 0.058 | -0.029 | -0.078 | -0.037 | 0.089 | 0.022 |
| empathic concern | ROG | IG | .302** | .211** | 0.034 | 0.101 | 0.011 | -0.007 | -0.012 | 0.118 | .245** |
| | IG | ROG | -.319** | -.183* | 0.085 | 0.057 | -0.003 | 0.090 | .222** | -.169* | -0.078 |
| | NOG | IG | .271** | .204* | 0.024 | 0.116 | 0.013 | 0.034 | -0.002 | 0.094 | .241** |
| | NOG | ROG | -.209** | -0.125 | 0.075 | 0.045 | -0.082 | 0.065 | .179* | -0.127 | -0.009 |
| | IG | NOG | -0.089 | -0.023 | 0.034 | 0.118 | 0.030 | 0.096 | 0.133 | -0.115 | 0.040 |
| | ROG | NOG | 0.107 | 0.102 | 0.096 | 0.101 | 0.004 | 0.079 | 0.038 | -0.010 | 0.105 |
| schadenfreude | ROG | IG | 0.084 | -0.036 | -0.018 | -0.126 | -.211** | -0.136 | 0.091 | .183* | -0.031 |
| | IG | ROG | .406** | .234** | -0.051 | 0.008 | -0.089 | -.158* | -0.139 | .236** | .181* |
| | NOG | IG | -0.045 | -0.084 | -0.076 | -.203* | -0.153 | -0.145 | 0.065 | 0.126 | -0.047 |
| | NOG | ROG | .355** | .181* | 0.029 | -0.016 | -0.071 | -0.116 | -0.109 | .173* | 0.165 |
| | IG | NOG | .182* | 0.051 | -0.012 | -0.122 | -.185* | -0.130 | -0.051 | .241** | 0.022 |
| | ROG | NOG | 0.035 | -0.086 | 0.037 | -0.110 | -.169* | -0.102 | .192* | 0.069 | -0.013 |
| pain intensity | derby game IG-OG | | .429** | .336** | 0.126 | -0.053 | -0.019 | -0.032 | -0.053 | .167* | .177* |
| | other game IG-OG | | .333** | .260** | 0.066 | 0.052 | .188* | 0.019 | -0.079 | .170* | .223** |
| | perpetrator IG-OG | | -0.093 | -0.056 | -0.094 | 0.049 | 0.063 | 0.034 | 0.123 | -0.067 | -0.046 |
| empathic concern | derby game IG-OG | | .573** | .403** | 0.049 | 0.033 | -0.060 | -0.060 | -.172* | .247** | .280** |
| | other game IG-OG | | .475** | .329** | 0.042 | 0.120 | 0.073 | 0.035 | -0.093 | .176* | .279** |
| | perpetrator IG-OG | | -.202* | -0.080 | -0.056 | 0.045 | 0.081 | 0.062 | 0.101 | -0.129 | -0.130 |
| schadenfreude | derby game IG-OG | | -.449** | -.260** | 0.072 | -0.069 | 0.040 | 0.104 | .203* | -.165* | -.234** |
| | other game IG-OG | | -.428** | -.261** | -0.078 | -0.099 | -0.013 | 0.006 | 0.140 | -0.103 | -.238** |
| | perpetrator IG-OG | | .170* | 0.113 | -0.033 | -0.123 | -0.151 | -0.092 | -.207** | .221** | 0.044 |

*Note.* ** $p < .000$; .** $p < .01$; .* $p < .05$; SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; IAT – Implicit associations test; IG – ingroup; ROG – rival outgroup; NOG – neutral outgroup

**Experiment 4: Discussion**

In contrast to our previous studies, in this study the expected intergroup bias was observed in all of the empathic ratings measured: pain intensity, congruent affect/empathic concern, incongruent affect/shadenfreude. Moreover, not only that intergroup bias emerged, but the effect sizes were medium to large: the difference in empathic responses varied depending on the victim identity to a notable degree. However, the intensity and the form of bias varied depending on the specific ratings.

The pain of an ingroup football player was seen as more intense regardless of the identity of the perpetrator. In other words, fouls committed against "us" were always perceived as more painful than fouls against "them". This points to "ingroup love" as the main form of creating bias because ingroup enhancement happened in clashes with both the rival and non-rival outgroups. Although in football competitions in the literal sense every participating team is considered a rival team, we have already argued why we consider Red Star and Partizan to be the only relevant rivals to each other, and other participating teams to be "neutral opponents". For painfulness ratings of the tackles, however, this distinction did not matter. The damage done against "us" was greater than the damage against "them" regardless of whether the neutral other or the direct rival had caused it; "their" pain was smaller than "ours" regardless of whether "we" were responsible for the damage or someone else was.

Higher-level empathy ratings paint a more nuanced picture. When asked about their own congruent feelings of empathic concern following the painful fouls, fans had significantly more empathic concern for their own team than for the rival team; they had even more empathic concern for their own player when the rival outgroup had hurt them, and even less empathic concern for the rival when the ingroup was responsible for the foul play. Who was responsible for the foul mattered for incongruent emotions too, as schadenfreude was the greater for tackles when the rival outgroup was hurt by the ingroup compared to tackles where a neutral team player was to blame. Expectedly, schadenfreude for the ingroup was low in both situations and did not differ with respect to the perpetrator identity. This pattern points to both "ingroup love" and "outgroup hate" as mechanisms of creating bias. The main effects of victim identity clearly point that the ingroup has a privileged position: we feel sorry for their pain by default, and feel no positive emotions after witnessing their pain. However, our empathic concern is even more amplified when the pain was inflicted by the rival compared to another team's player. In contrast, the empathic concern for the outgroup pain is even more minimised when the ingroup had inflicted the pain, and incongruent (and normatively inappropriate) positive emotions emerge.

We observed biased responding when the victim of the foul was a member of a neutral team as well: the pain was less intense and fans had less empathic concern and felt more schadenfreude if the ingroup had committed the foul. Although this comparison does not directly contrast the empathic responses to "us" versus "them", it clearly points to the utility of biased responding. The competitive interactions with the neutral outgroups are only competitive in name, i.e. the outcome of the game had almost certainly been in favour of the rival outgroups; however, winning matters regardless, and this result points that empathic assessment in any competitive interaction is seen through the lens of group identity and biased in favour of the ingroup. However, the effect sizes of group identity were notably smaller when the neutral outgroup was on the receiving end of the tackle compared to ingroups and rival outgroups.

Empathic ratings in this experiment, in contrast to previous experiments, were reliably related to fan identity measures and social dominance orientation, as expected, but not to empathic personality traits. In our opinion, this result indicates that the assessment was perceived as a group-identity based situation and was made on behalf of the group. The relationship of biased pain intensity ratings, empathic concern and schadenfreude, with fan identity was rather strong ($r$ = .33-

.58). Moreover, highly identified participants tended to react more strongly to the ingroup pain especially in ingroup-rival outgroup interactions, pointing to the importance of strength of identification for empathy in group settings and both ingroup love and outgroup hate as mechanisms of creating and maintaining the relative ingroup advantage.

Taking into account the ecological validity of the stimuli, compared to physical pain studies we reviewed, this study was conceptually most similar to studies using videos of painful shoulder movement of racial ingroup and outgroups. In these studies, the stimuli were depicting real-world physically painful situations and it was plausible to expect the participants could find themselves in a position to make such empathic assessments (or they already did, as in some studies the participants were medical professionals).

Two of these studies that did find bias (measuring empathic concern) were conducted on student samples; two that did not (one measuring empathic concern, one assessing pain intensity, severity and attribution) were conducted on (future) medical professionals. Although the context of the task was not group-identity relevant, i.e. participants were asked to assess the pain of an individual in medical settings, medical professionals might have been more resistant to group identity signals or have been more able/motivated/knowledgeable on how to discard them. In other words, studies using community samples might have included individuals who did observe the assessment as identity-relevant (not necessarily in terms of high ingroup identification or having negative attitudes towards the racial outgroup, but for example as reflecting stereotypes about the outgroup's relationship to pain, which e.g. the Trawaler et al studies point to).

In our study, football fans were asked to assess a situation they encounter every time they watch a football game, i.e. to provide an assessment they spontaneously make in their everyday life; however, the context in which they make those assessments is explicitly group-identity based. In other words, they were explicitly pushed into an ingroup versus outgroup situation and provided the opportunity to validate their group identity. As we have demonstrated, in those situations ingroup bias in empathic responses readily emerged.

## General Discussion

We will first briefly summarise the similarities, differences, and main results of all experiments and then attempt to embed them within contemporary socio-psychological accounts of intergroup bias and theoretical models of empathy. We will then review the concept of ingroup empathy bias as reflected in current scientific practice and discuss it in the context of broader meta-scientific issues of conceptual systematization and clarity. We will identify the causes and consequences of (mis)communication of scientific findings both within and outside the field. Lastly, we will provide recommendations regarding studying intergroup empathy bias specifically and doing and communicating science more generally.

Table 27
*Experiments 1-4: summary*

| Exp. N | Sample | Design | Stimuli | Dependent measures | Results | Bias, Traits and ID |
|---|---|---|---|---|---|---|
| Experiment 1 | 147 | 3 (group: IG, ROG, NOG) x2 (painfulness) | images of painful (needle) and neutral (q-tips) events | pain intensity, self-unpleasantness | no intergroup empathy bias | none |
| Experiment 2 | 140 | 3 (group: IG, ROG, NOG) x2 (painfulness) | images of painful and neutral facial expressions | pain intensity, self-unpleasantness | no intergroup empathy bias | IRI personal distress (weakly) |
| Experiment 3 | 97 | 3 (group: IG, ROG, NOG) x2 (salience) | video recordings of models retelling painful everyday misfortunes | pain intensity, self-unpleasantness, schadenfreude, empathic concern, perspective taking | no intergroup empathy bias | SDO (weakly) |
| Experiment 4 | 158 | 2 (victim: IG, ROG) x 2 (perpetrator: IG/ROG, NOG) 1 (NOG victim) x2 (perpetrator: IG vs ROG | images of painful fouls from real football games | pain intensity, empathic concern, schadenfreude | - pain intensity: ingroup favouritism - empathic concern, schadenfreude: ingroup favouritism + outgroup derogation | ID (strongly) and SDO (weakly) |

*Note*. IG – ingroup, ROG – rival outgroup, NOG – neutral outgroup, SDO – social dominance orientation; ID – identification, IRI – Interpersonal reactivity index

In Experiment 1 and Experiment 2, to investigate intergroup empathy bias for physical pain, we employed two experimental paradigms typically used in previous studies: we presented the participants with photos of painful versus neutral events happening to faces (with a neutral facial expression) (Experiment 1) and painful vs neutral facial expressions (Experiment 2) of ingroups, rival outgroups, and neutral outgroups. As opposed to photos used in the first two experiments, in Experiment 3, we used dynamic stimuli – video recordings of ingroups, rival outgroups, and neutral

outgroups retelling a painful event that had ostensibly happened to them. To manipulate identity salience, participants were told it had happened while the actor had been doing something fan identity-related or while doing a comparable everyday task. To further enhance ecological validity and identity relevance, in Experiment 4 photographs of painful illegal tackles i.e. fouls from real football games with ingroups, rival outgroups, and neutral outgroups as both victims and perpetrators were used as empathy-eliciting stimuli.

In all experiments, participants reported on their fan identity only after rating the stimuli and were provided no specific context for the task except for a vague explanation in the instruction that the goals of the study were related to physical pain assessment. Within the tasks, the physical pain in Experiments 1 and 2 was either symbolically or directly communicated, but decontextualised, i.e. no information on its causes or consequences was provided. In Experiment 3, physical pain was placed in an everyday context and described (as compared to visually presented) as a personal event; the cause of physical pain was a physical accident that could have happened to anyone. Although the fan identity of the person in pain was made salient in 50% of the events, the events themselves were not explicitly group-based and were inconsequential to the group identity of the person, i.e. they were individual misfortunes. In Experiment 4, physical pain was placed in the context of intergroup competition, i.e. the core of the division between us and them, and it was thus related to the success of the group in that context.

In Experiments 1-3 we detected no ingroup bias in empathic responses, but in Experiment 4 the expected bias in favour of the ingroup was reliably present and substantial in all empathic measures we assessed: pain intensity, empathic concern, and Schadenfreude. We argue that this pattern fits well within the social identity approach to bias as a contextually defined meaningful response to a social situation. For empathy bias, these attributes imply that to observe bias, the empathy-eliciting situation needs to be group-identity relevant and the biased response needs to be functional in maintaining that group identity. The empathy-eliciting tasks in four experiments differed in how relevant and functional the biased response was. Specifically, we will discuss the differences in cognitive and motivational aspects of the experimental situations with respect to their group-based character. In addition, although our results could be seen to go in line with the motivated empathy account (Zaki, 2014) and other views on empathy as a strategic response (e.g. Weisz & Cikara, 2021), we will point to several discrepancies in the way those views define, measure and discuss empathy and how we interpret the results of our studies in the context of motivated ingroup biases.

## Intergroup Empathy Bias and Social Identity

Social identity approach (Hornsey, 2008) – Social identity theory and social categorization theory – posits that human interactions represent a spectrum – from purely interpersonal to purely intergroup – and the shift back and forth from one part of the spectrum to another changes the way people see themselves and other people. Social identity is seen as a part of individual identity derived from the social categories we belong to and the emotional and evaluative consequences of such membership (Tajfel & Turner, 2004). We favour our own group over the others because we are motivated by a desire to have a positive self-concept, or to achieve, maintain or protect a positive distinctiveness between our own group and relevant outgroups.

Once we recognise ourselves as members of a certain group, and attach a certain importance to that specific group membership, the preconditions for social identification are met (Tajfel, 1982). The former – the process of social categorization – is defined as a cognitive shift: it enables us to classify and order the social environment and thus participate in and navigate through the social

world, but it also provides a basis for self-reference in social terms (Tajfel & Turner, 2004). Social identification implies not only classifying ourselves as group members but also valuing being a member – hence it changes the way we perceive and evaluate ourselves and other people and changes our motivations to interact with other people – both become based on group membership. Once our social identity is salient, we start acting as group members and not individuals, and this shapes the way we see, think, and act in specific social contexts.

However, social categorization is not seen as a blind and automatic process, but as a highly flexible and context-dependent one; this enables us to mobilise different group identities for different purposes (or not mobilise them at all) (Tajfel, 1982). Flexibility and context-dependency imply that social categorization represents a functional response, i.e. that it shapes our psychological landscape only when it helps us navigate the social environment. What follows is that perceivable social categories will not necessarily shape or influence our responses if they do not contribute to our adaptive response to the situation at hand (Tajfel, 1982).

Moreover, even if in certain contexts we perceive the world through specific social categories we value, our response to those situations is not expected to be algorithmic, but flexible and responsive to various objective and subjective characteristics of social context (Tajfel, 1982; Tajfel & Turner, 2004).

We argue that our experiments critically differ with respect to the cognitive and motivational functionality of a group-based response. From a cognitive standpoint, the experiments differ with respect to how salient the group identity is in terms of accessibility and fit (Oakes, Turner & Haslam, 1991), or, in other words, how meaningfully the experimental tasks could be navigated by employing group membership categories. From a motivational standpoint, the experiments differ with respect to how threatening they are for our group identity and therefore how functional a group-based response may be in a specific situation (Ellemers, Spears, & Doosje, 2002).

### *A Cognitive Account: Accessibility x Fit*

In an attempt to specify the conditions under which we start observing the social world from the perspective of a specific social identity we endorse, Oakes and collaborators (Oakes, Turner, & Haslam, 1991; Haslam, Oakes, McGarty, Turner, & Onorato, 1995) have introduced the concept of group identity salience as crucial for understanding group-based self-definition. Salience, however, is not an internal quality of the stimulus that influences our attention regardless of its social meaning, but a functional response important for navigating the social situation.

Social categorization is a context-sensitive process: whether a certain social category will become salient depends on the interaction of accessibility and fit (Oakes, Turner, & Haslam, 1991; Hasslam, Oakes, McGarty, Turner, & Onorato, 1995). Accessibility refers to the readiness of a given category to become activated in a person's perceptual field; therefore, accessibility is sensitive to individual and contextual influences. Fit, however, depends on the relationship to reality, or in other words, a category is as salient as it is useful for representing a particular social situation. Fit has two aspects: comparative and normative. Comparative fit refers to the perceived meta-contrast of the social categories in a specific situation – what is the "F statistic" of a social situation, and how much the perceived similarities and differences vary within a group compared to between groups. A social category must be perceived to correlate systematically with the observed similarities and differences. The normative aspect of fit refers to the content of the actions and implies a match within a stereotypical representation of a social category (Oakes, Turner, & Haslam, 1991).

We applied this framework to discuss the pattern of results in Experiments 1-4. We designed all experiments so that the social categories of fan ingroup, rival outgroup, and distant outgroup were readily accessible to the perceiver. Models or players were wearing highly recognizable jerseys, and the jersey was the only visually displayable social indicator available (we cannot exclude individual stereotypes about faces or common knowledge or belief about a specific player, but in prominent social terms, all models were young white males). As for individual differences influencing fit, participants were highly identified individuals. However, we believe the fit to be relatively more accessible in Experiment 4 as the context of assessment has not one but multiple signals pointing to the group aspect of the situation: the interaction between players whose consequences participants rate happens on a football field, and is clearly something that had happened and happens regularly within an institutionalised rivalry that represents the basis of participants' group identity – it should be therefore more accessible to a fan than an image of an unknown person with a painful expression wearing their team's shirt.

However, accessibility is not sufficient, and Experiments clearly vary in terms of both comparative and normative fit – and both are low in Experiments 1-3. Comparative and normative fit are conceptually distinct, yet closely intertwined in practice in most cases, and it is, therefore, difficult to separate them in analysing a specific case.

As for the comparative fit, there are several arguments. To our knowledge, there is no specific stereotype or belief related to either of the fan groups in our stimuli that would help the participant provide an answer on how painful a stimulus is. This is not the case with racial categories: studies in the United States (Trawalter, Hoffman, & Waytz, 2012; Hoffman, Trawalter, Axt, & Oliver, 2016) identified a number of false beliefs about biological differences between racial groups which allegedly make Black people less sensitive to pain, and these beliefs were shown to shape the painfulness estimates. In addition, there were no other social characteristics of the assessment itself except for the colour of the jersey, and the group identity was not related to the situation depicted in the stimuli in any way that would help explain it or help the participant respond to it – neither to the pain expressed or inferred in the images nor the pain described in the video events. Although a part of the painful events in Experiment 3 was designed to make the fan identity of the model salient, the situation in which the pain happened was not related to the group-based rivalry in any way. However, we want to point out that had our salience manipulation been successful, it could also be interpreted as an increase in comparative (and normative) fit.

As for the normative fit, general social norms related to others' pain prohibit gloating and prescribe empathic reactions, and the assessment was not framed as social in any way that would make breaking these norms justifiable; we believe they were employed when responding to stimuli and that participants did not respond as group members. This response varied in intensity depending on the empathic potential of stimuli, but not on social dimensions.

However, in Experiment 4, the stimuli depicted an event easily and efficiently described in social terms, in both comparative and normative terms. The participants rated their responses to painful events happening to ingroup and outgroup targets within an institutionalised competition which forms the basis of the division into us and them. Hence, the social meaning of the situation prompted responding on behalf of the social self. The event presented in the stimuli was best described in terms of fan identity and had a specific social meaning. In addition, the pain was inflicted within a normatively defined identity-related interaction readily available to describe (and legitimise) in group terms.

To become salient, a social category must be both accessible and fit to explain the situation. We argue that only in Experiment 4 did fan identity as a category fulfill both criteria. That is, it was salient enough to steer the participants toward social categorization, as it was both available and appropriate for the context. In other words, only in Experiment 4 did participants respond on the basis of their fan identity, crucially because of the higher comparative and normative fit of social

responding, compared to Experiments 1-3. In Experiments 1-3, the social identity was available but not fit for the context of assessment.

In this analysis of salience, social categories are not observed as fixed but dynamic entities and the process of social categorization is seen as sensitive to specific social contexts and varying according to specific relations as well as specific behaviours in question (e.g. Andersson, 2010). Applied to our case, this account does predict the possibility of different patterns of biases to emerge in different responses, as was the case in Experiment 4 where we detected ingroup favouring tendencies for pain intensity but outgroup derogating as well in higher-level empathic ratings. In addition, it does allow for the influence of individual difference variables on social identity salience, which can be inferred from the significant correlation of identification and SDO with bias. However, this accessibility x fit account does not provide us with specific expectations regarding different ratings, therefore different patterns of bias could fit within the account as well.

Despite the variability in definitions of empathy, based on the review study we infer that the emotional quality of empathy is almost universally present in concept and/or operational definitions. If we observe empathic reactions as emotional reactions, the different patterns of bias could be interpreted through intergroup emotions theory as well. Intergroup emotions theory focuses on the emotional consequences of self-categorization in group terms (Mackie, Smith & Ray, 2008). When a certain social identity is salient, emotions are felt on behalf of the group, and people react in a tailored manner to specific events and objects that affect the group they identify with. Self-categorization is suggested to influence emotional reactions because when social identity is salient, events are being appraised on behalf of the group, and because the person engages in self-stereotyping that shapes their responses in the expected direction. Intergroup emotions are important because of their consequences, the most important being their influence on imagined and actual behaviour, in a fine-tuned manner. In other words, emotions have action-motivating potential.

In Experiment 4, participants assessed the pain intensity of the foul, as well as their empathic concern and Schadenfreude. If we treat these three ratings as different emotional reactions – and we argue for this position – it follows from the intergroup emotions theory that different patterns of bias could emerge with respect to their implications to group identity and that they would be systematically related to social identification. Pain intensity ratings represent an (emotional) stimulus intensity rating, and empathic concern and schadenfreude are considered as higher-level empathic ratings, congruent and incongruent, respectively, and represent an other-directed response. It is easy to post hoc justify why only the ingroup favouring bias in pain intensity estimates should benefit the group position, but both ingroup favouring and outgroup derogating biases in higher-level ratings could be seen as useful for the group. For example, it could happen that for emotion intensity ratings additional conditions are needed to differentiate between outgroups, one of the obvious ones being competition. However, the truth is that Intergroup emotions theory also does not offer specific expectations about empathic ratings or any other emotions. Other than making a general preposition that different intergroup emotions influence different patterns of thought and action in specific ways with group benefit in mind, it leaves the specifics to empirical studies.

In summary, the accessibility x fit account provides us with a coherent explanation of why the bias was only registered in Experiment 4, and a general expectation of variability of biased responses with respect to numerous individual and structural variables. Neither this account nor intergroup emotions theory provide a basis for spelling out specific expectations about different empathic ratings, hence our suggested explanations are post hoc and should be investigated in further studies. However, both point to flexibility and situational specificity of our responses to ingroups and outgroups, from the perspective of the group and with group benefit in mind. In other words, our responses are functional and motivated by group benefit. We now turn to the analysis of those motivational aspects of the experimental tasks the participants were exposed to.

### *A Motivational Account: Identification x Threat*

Biases in cognition, affect and behaviours represent the means to create and maintain a positive view of those aspects of the self which are rooted in group membership. In spite of being functional and wide-reaching, the social identity approach does not predict biases to be universally observed, but to be sensitive to socio-historical circumstances, reality constraints, and contents of group norms (Hornsey, 2008).

Ellemers, Spears, and Doosje (2002) argue that according to the social identity perspective, identity concerns stemming from group membership are determined by group commitment (i.e. strength of identification) and important features of social context that shape our responses in group situations – or more specifically, threat. They develop a taxonomy of situations reflecting different identity concerns that arise due to differences in group commitment of the individual (low vs high) and the type of threat that can be identified in context (no threat, personal threat that stems from our associations with the group, and threats to the group itself). According to this view, different identity concerns – and different underlying motivations – influence our perception, affect, and behaviour in a different manner.

The crucial idea we rely on when interpreting our results based on this taxonomy is that social identity is a perceiver factor implicating different aspects of the self. Each person brings certain individual traits and certain valued social identities into every social interaction. On the other hand, interactions are characterised by various social contextual factors that enhance or diminish the meaningfulness of personal as well as social identities (Ellemers, Spears, & Doosje, 2002). Together, they shape our responses in specific situations in a meaningful way. However, this implies that, no matter how salient the social identity is, it can influence the same assessments strongly in some contexts but not in others, depending on the usefulness of such assessment.

In terms of threat x commitment taxonomy, all our studies belong to the high commitment quadrants, as the participants were pre-selected for the strength of identification. In Experiments 1-3, no threat to the participants' personal or group identity was implicated in the tasks, hence they belong to the no threat-high commitment quadrant. In contrast, in Experiment 4, groups are placed in a competitive context and the comparative success or failure threatens the value of the group in terms of status.

In our interpretation, we studied biased perception (pain intensity) and biased affect (self-unpleasantness, schadenfreude, empathic concern, perspective-taking) in all experiments.

According to this taxonomy, highly committed group members in non-threatening situations are concerned with creating a distinguished identity (in groups yet to become) or expressing and affirming their identity (in groups with clear-cut identities). As the self-relevance of the group is high, the expected effects range from perceptual to affective and behavioural. However, as the situation is not threatening, the effects are expected to be limited to ingroup enhancement in perceptual domains, although in behaviour the motivation for differentiating between ingroups and outgroups could lead to outgroup derogatory consequences. In domains of affect, this taxonomy primarily discusses the self-esteem hypothesis by stressing that high commitment will probably implicate collective self-esteem.

In contrast, in group-threatening situations, biases in perception, affect, and behaviour are instrumentally aimed at group differentiation and group affirmation. Group threats may vary, and have different perceptual, affective, and behavioural consequences; in other words, depending on contextual and structural differences, the bias may take a different form.

As in our experiments 1-3 there were no individual or group threats implicated, and pain represents a negative and disturbing stimulus difficult to use to achieve positive distinctiveness, the

absence of bias in pain intensity ratings in these experiments is in line with both theoretical and empirical prediction of Social identity theory.

For self-unpleasantness (Exp 1-3) and positive other-oriented empathic reactions we assessed (empathic concern and perspective taking, Exp 3), the predictions are not as clear. The authors of the taxonomy rely mainly on the infamous self-esteem hypothesis; along with the increased self-relevance of the group, it can be expected for the pain of the ingroup to be more disturbing to the self compared to the pain of the outgroup. However, once again, pain is a universally negative sensory and emotional experience; in line with our motivated perspective on intergroup bias, we fail to see how would biased emotional reactions to unambiguous negative stimuli in non-threatening and above all socially meaningless circumstances help a person express and maintain a positive identity of the group they belong to.

Although the stimuli were decontextualised and provided no opportunity for group identity enhancement in Experiments 1 and 2, it can be argued that in Experiment 3 ingroup bias could be expected to emerge. The stimuli were more complex and ambiguous with regards to signal clarity (Matsumoto, 2002) or, in other words, they provided more opportunity for motivational interpretation compared to previous experiments; in addition, we measured empathic concern and compassion, and being biased in those reactions can in principle be interpreted as making a positive distinction between the ingroup and the outgroup (especially if we have in mind that extension of "empathy" is often listed as an initial form of discrimination between us and them, see Hewstone, Rubin & Willis, 2002). In other words, finding bias, especially in higher-level empathy ratings, would fit into social identity theory predictions as well. However, we registered no bias in either pain intensity or higher-level empathy ratings. Our results indicate that in a threat-neutral context, people do not display ingroup bias in several empathic ratings for physical pain. Compared to previous similar studies, this result gives physically painful incidents a special status, as the number of misfortunes studies finding ingroup bias is substantial.

For incongruent empathic reactions i.e. schadenfreude the predictions are clear – in the absence of threat and explicit competitive motivation, the negative asymmetry between groups is not expected to emerge.

We emphasised several times that pain is a negative stimulus. Bearing in mind the positive-negative asymmetry in studies of ingroup bias (i.e. the fact that bias is often found for positive but not negative attributes) (Hewstone, Rubin & Willis, 2002), as well as the absence of personal or group threat in our experimental design, we believe our results fit within the predictions of social identity approach. In other words, we believe the motivational aspects of bias to be absent in our experiments, hence the bias did not emerge. Or, in terms of social categorization theory, pain assessment in non-threatening contexts for the highly identified predominates either personal or general human identity (bearing in mind the weak correlations between trait measures and empathic ratings in Experiments 1-3, there is tentative evidence in favour of the former).

In contrast to Experiments 1-3, the empathizing task in Experiment 4 was embedded in an explicitly competitive interaction between groups. In this context, both perceptual bias (ingroup favouring pain intensity assessment) and affective biases (increased empathic concern for the ingroup and increased schadenfreude for the outgroup, especially in most competitive situations, i.e. when clashing with the rival outgroup) are clearly functional in terms of maintaining group status. In our experiment, the perceptual bias took the form of ingroup favouritism, but in the emotional reactivity biases, both ingroup favouritism and outgroup derogation can be implied. "This foul must really hurt our player, but those playing against us are faking it" is evidently aimed at affirming ingroup superiority. However, it seems that rivalry further differentiates the emotional responses to ingroup pain, implying some form of outgroup derogation. If we hypothesise higher or different motivation in comparisons with rival versus neutral outgroup, our results indicate that emotional reactions (empathic concern and schadenfreude) are modified by these differences in motivation

and perceptive estimates (physical pain intensity) are exempt. Our results once again support the idea that pain intensity ratings have a special status. We observe them to be susceptible only to ingroup favouring but not outgroup derogating forms of bias, which asks for further investigation. However, once again we want to stress that empirical evidence for ingroup bias in pain intensity ratings only in ingroup favouring form and also in both forms could fit into social identity theory predictions.

### *Conclusions*

The results of our experiments generally go in line with the predictions of the social identity framework – social identity theory and social categorization theory, as instantiated in specific cognitive and motivational frameworks of analysis. According to our studies, for ingroup bias in pain empathy ratings to emerge, one way is to make the context of assessment related to group identity concerns i.e. to make it motivationally relevant; or in cognitive terms to make the group identity salient, meaning both highly accessible and contextually fit.

We observed no bias in the decontextualised assessment of pain intensity and self-unpleasantness via ratings, under neutral conditions. This finding has important implications for behavioural validation in neuroscience studies using the same paradigms.

Whether painful stimuli in general are less sensitive to bias in non-threatening conditions of group identity accessibility should be validated in further studies. In Experiment 3 we did not register bias, implying that group identity does not influence our perceptions and reactions to other's physically painful misfortunes happening outside of group-related context. As there are a number of studies on miscellaneous misfortunes detecting bias, further studies are needed to clarify whether physical pain as a stimulus has special properties and its assessment and response are less sensitive or differently sensitive to group-based influences, or we were unable to detect bias due to our choice of reactions measured or measure sensitivity. Either way, the explanation should be drawn from the broader accounts of bias and in relationship with other studies on bias across groups, contexts, methods, and measurements.

Social identity theory and the social identity approach in general, have been criticised on several grounds (for a historical review see Hornsey, 2008). For the present discussion, the most relevant critique is that the theory has become too broad to be falsifiable. Indeed, we have also pointed to several results that could be interpreted within the cognitive or motivational account, but so could several possible different ones. However, if we understand the social identity approach as a meta-theory about the social self, we imply that its basic principles can be applied broadly and are broadly evaluated, by conceptual and theoretical analysis of multiple empirical findings, sometimes from superficially different research lines. In Meehl's terms, social identity theory is a "soft theory" that is in principle unfalsifiable by a single significant or non-significant result, as it makes directional but not point predictions (Meehl, 1990).

Therefore, although we acknowledge the underspecified nature of social identity theory and its auxiliary theories in particular, being able to incorporate alternative findings is not sufficient to nullify its benefits as an overarching framework that explains when and why we can expect group-based self to guide behaviour. However, we stress the need for conceptual clarification and classification, especially in so-called auxiliary theories (Meehl, 1990) with respect to important differences in theoretical and operational definitions of group identity-influenced phenomena and important structural variables defining the social environment.

In conclusion, intergroup bias in empathic responses to painful events emerges when the social identity is both accessible and provides an appropriate framework for describing and responding to a social situation. This biased response is shaped by individual and social variables influencing our motivation to see, feel, and act in a certain manner.

## Intergroup Empathy Bias and General Models of Empathy

In the systematic review of intergroup empathy bias research, two broad theoretical accounts of empathy were typically referred to: the Perception-Action model (PAM) of empathy (Preston & DeWaal, 2002), and the Motivated empathy account (Zaki, 2014). Both models can incorporate intergroup empathy bias as an empirical phenomenon. However, neither of these models makes explicit connections to any socio-psychological theory of bias.

In the PAM, empathy is defined broadly, as any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation (Preston & De Waal, 2002). Importantly, the PAM classifies empathy as one of the processes relying on a broadly conceptualised perception-action mechanism, an idea that perception and action share a common code of representation in the brain. This mechanism is viewed as a general feature of the nervous system that is adaptive because it facilitates the mother-offspring bond, as well as group living and social responding outside of the parent-offspring relationship. Another important feature of the model is that it assumes state representation resulting from the attended perception of an object automatically generates the associated autonomic and somatic responses unless inhibited. Therefore, although it broadly defines empathic phenomena (and explicitly postulates that co-activation of representations employing perception-action mechanism is responsible for relatedness and inseparability of empathic phenomena such as emotional contagion, sympathy, empathy, cognitive empathy, and prosocial behaviours), the PAM restricts the term empathy for co-representational and automatic state matching at a representational level. This definition substantially matches the implicit definition of empathy in neuroscience studies – neural co-activation.

However, state matching is not expected to necessarily result in a visible behaviour. Moreover, shared representations are believed to change with experience; they are expected to be richer for familiar and similar objects, sensitive to past events and implicit and explicit learning, and differentially activated depending on cue salience. In addition, the PAM explicitly states that along with matching state representations, various other representations about the object and the situation also become activated, and their interaction determines the appropriate response – for example, the appropriate response could be both prosocial and antisocial behaviour, depending on whether we are interacting with a friend or an adversary.

The PAM therefore allows, and we would argue that it also predicts for social categorization to influence empathy via multiple mechanisms: for example, through perceived ingroup similarity, shared norms of thinking and feeling in stable social groups with a common history, through changing expectations, through contextual influences, etc. What the PAM suggests is that these influences can shape the richness of the representations, and the probability of their activation which is expected to depend on cue salience, but whether these representations will be acted upon depends on influences external to representational co-activation. In summary, the PAM is a broad evolutionary account of empathy that restricts the meaning of empathy to automatic co-representation of states, and specific predictions regarding ingroup bias lie outside its domain. In general terms, ingroup bias can be expected but not necessarily manifested in visible behaviour. Therefore, both finding and not finding bias in our studies could be interpreted within the PAM.

However, considering the meta-scope of this account, the discussion about social identity theory applies here as well.

The PAM's view of empathy as *automatic, unless inhibited,* is problematised by many authors as failing to account for many aspects of empathic experience or its absence in everyday life. Specifically, by emphasizing the stimulus-driven feature of empathy, its motivational basis and proactive regulation are sidelined (Ainslie & Monterosso, 2002; Bandura; 2002). Bandura argues that the PAM is "predicting more than has ever been observed" and that such an automatic empathy mechanism would take an enormous emotional toll on the perceiver and paralyze everyday interaction (Bandura, 2002). Automaticity of representations aside, we argue that its silence about the motivational influences on empathy makes the PAM unsuitable to interpret the specific results of our Experiments in this framework.

The Motivated empathy account on the other hand views empathy as a functional response and elaborates on its motivational basis. Explicitly, empathy is defined as sharing and understanding the affective states of others, and differentiates between experience sharing, mentalizing, and mind perception as components of empathy, the latter considered to be a precondition to empathic engagement (Zaki, 2014). The key feature of this model is that it questions the hypothesis embedded into contemporary models that empathy (or more precisely, experience sharing) is an automatic mechanism for catching and matching the states of others. Instead, empathy is observed as a motivated phenomenon: individuals can be driven to engage with (i.e. approach) or disengage from (i.e. avoid) other people's emotions.

In the Motivated empathy account, the intergroup empathy bias[16] is seen as evidence that empathy is a motivated process and subsumed under "contextual effects on empathy". In addition, along with offspring care, ingroup identification and outgroup exclusion were listed as ultimate sources of empathic motives; both identification and exclusion are suggested to exist because they had provided an adaptive advantage during the course of human evolution and are therefore deeply ingrained in our response patterns. Those motives affect empathic processing through differential information processing and emotion modulation, by employing strategies such as situation selection, attention modulation, and differential appraisal.

According to this model, one of the reasons people might be motivated to avoid empathy is because they are motivated to avoid pain caused by empathizing. The other two empathy avoidance motives are avoiding material costs of empathy (e.g. being asked for a charitable donation) and avoiding interference with competition.

Our studies did not implicate any material cost of empathizing for the participants, hence the latter motive is not relevant. However, our results can be interpreted as evidence for the interference with competition as a motive to avoid empathy, as the only experiment where intergroup bias in empathic responses emerged was the one where the pain was evaluated within explicitly competitive between-group interactions. The relationship of our studies with avoiding pain as a motive is less clear: if we take as given that seeing a signal of incoming pain (Exp1), a painful face (Exp2), or hearing about pain (Exp3) is sufficient to elicit an empathic reaction in participants strong enough to be modulated (because empathy is costly and there is not enough of it for everybody), then our results indicate there is no ingroup bias in several explicit empathy ratings. However, we have already expressed our doubts about the strength of the empathic reaction caused by stimuli in Experiment 1 and Experiment 2. If we take the results of Experiment 3 as reliable, they indicate that avoiding pain is not a motive for empathizing differently with ingroups and outgroups in pain in baseline circumstances, i.e. when no personal or group benefit would follow biased responding. However, the authors of the model refer to ingroup bias registered in neural responses as evidence for the motivated nature of empathy. As we have previously highlighted,

[16] Under the umbrella term intergroup conflict, the author discusses studies of various empathic phenomena that observed ingroup bias dominantly via psychophysiological responses.

neuroscience studies of empathy operationally (and implicitly theoretically) equate empathic responses to shared representations, supposed to happen automatically; as neural co-activation is weaker for outgroups, this is suggested to reflect the motivational influences on empathy stemming from group membership. As we have only assessed self-reported empathic responses, we are not certain if our results could be interpreted as speaking against pain avoidance – it is probable that they would only be interpreted as a lack of bias in explicit ratings.

The conceptual problems in empathy research in general are visible in the motivated empathy account as well. Empathy is defined broadly, as both experience sharing and mentalizing. The implicit assumption about automaticity is mainly drawn from evolutionary, developmental, and neuroscientific studies and theoretical considerations of empathy, and studies listed as evidence for motivational modulation of empathy range from diminished neural resonance to outgroups compared to ingroups to increases in antisocial behaviour to outgroups in a football competition. Although we generally agree with the motivated empathy account, we argue that to understand the nature of ingroup empathy bias, we need to consider the kinds of empathic reactions that show or do not show bias depending on the context in which the empathizing occurs. As it is hopefully clear from our discussion above, no socio-psychological account of bias considers it a blindly automatic phenomenon, but a functional response happening in a social context.

We now turn to summarise the review and empirical findings of this thesis, elaborate on conceptual issues in empathy research, and discuss how they reflect on the study of intergroup bias in empathy, both in terms of our understanding and knowledge about the phenomenon and in terms of the impact of that knowledge on further scientific research and public impact alike.

## Intergroup Empathy Bias? Conceptual and Meta-theoretical Reflections

In the systematic review of intergroup empathy research, we identified considerable conceptual heterogeneity, which was expected for a phenomenon under such an umbrella term. This heterogeneity is reflected in our studies as well, as two studies were specifically designed to replicate a selection of frequently used paradigms to elicit empathy, and two were to a different degree inspired by the frequently used empathy-eliciting tasks and designed to examine the effects of contextualisation of pain. We observed no bias when participants were asked for decontextualised pain assessment. When pain was assessed in an ecologically valid context, we observed bias only when the context was directly related to the participant's group identity. We interpreted the pattern of results in the framework of social identity theory and concluded that intergroup bias in empathic responses to painful events emerges when the social identity is both accessible and provides an appropriate framework for describing and responding to a social situation; otherwise, it should not be expected to manifest. This biased response is shaped by individual (identification) and social variables (group-based threat) affecting our motivation to see, feel, and act in a certain manner. As we observed both ingroup favouritism and outgroup derogation depending on the specific empathic response, we leave the question of ingroup love vs. outgroup hate unresolved regarding empathy bias. Put differently, we can reliably claim only that the emergence of intergroup bias is highly sensitive to individual and social contextual differences as well as specific measures of empathy.

We argue this collection of "it depends" in our summary is indicative of the conceptual disorder existing in empathy research, the underspecificity of theories in social psychology, and the absence of theoretical communication between those two areas of study. In turn, these issues reflect the disbalance of the epistemic triangle in psychology in favour of empirical investigation at the expense of theoretical (Machado, Lourenço, & Silva, 2000). Moreover, we argue this theoretical

state-of-the-art has harmful consequences for our attempts to understand empathic phenomena and for the public image of psychological science.

We find Meehl's analysis appropriately titled *Why summaries of research on psychological theories are often uninterpretable* (Meehl, 1990) highly relevant for understanding the body of research we have reviewed and produced; therefore, we are going to use it as a framework for discussing the obstacles we identified preventing us to reach more substantial conclusions. However, as Meehl's paper contains highly complex and nuanced insights, we cannot do justice to it in a short summary, therefore we recommend it in full to both interested and non-interested readers.

Thirty-three years ago Meehl argued that in so-called "soft" areas of psychology (social psychology, personality psychology, etc.), null hypothesis testing is an inappropriate way to validate a theory. In spite of that, this is still a common practice in research and meta-analyses alike. Most theories in psychology are "weak" in the sense that they at best can make predictions about directional difference or association without specifying its size, as contrasted to point predictions made by "strong theories". Although refuting the null hypothesis and accepting the alternative hypothesis are separate conclusions different in kind (the former is statistical and the latter a theoretical conclusion), in the scientific language they are often equated.

Meehl argues this logical slippage has a disproportionate effect on the falsifiability of "weak" theories and leads to many "mixed" findings in reviews of many theoretical accounts in psychology. The disproportion comes from the influence of "obfuscating factors" in measuring complex multicausal phenomena, whose effects are, to directly cite Meehl, *"usually (1) sizeable, (2) opposed, (3) variable, and (4) unknown. The net epistemic effect of these ten obfuscating influences is that the usual research literature review is well-nigh uninterpretable."* (Meehl, 1990, p. 197) In other words, these influences are expected to vary from domain to domain, from experiment to experiment, instrument to instrument, they are typically unassessed and may be impossible to estimate in principle. However, we continue to behave as though a single study using a specific operationalization of a complex phenomenon measured in a single context on a specific sample confirms or disputes a theoretical statement in general by confirming or refuting the null hypothesis (Yarkoni, 2022). Therefore, if we summarise those single studies all saying to measure the same phenomenon, all we can possibly get in a review is a bag of mixed results.

Some of the influences Meehl listed have been recognised (some under different names), and discussed and attempts have been made to address them during the last decade, such as (1) inadequate power, (2) experimenter error, (3) not publishing pilot studies and (4) selective bias in submitting reports (i.e. "file-drawer effect"), selective editorial bias in publishing positive findings (5), (6) unaddressed questionable validity of instruments. Most of these, according to Meehl, "make bad theories look good"[17], i.e. increase the probability of false positive findings interpreted as speaking in favour of a theory (as theories are "weak", many results can fit in their predictions and false positives are easy to find under these influences). Attempts to mitigate these influences are reflected in many formal initiatives (the most famous being the Reproducibility project, (Open Science Collaboration, 2015) and informal behavioural patterns (e.g. more frequent requests and inclusion of unpublished datasets in meta-analyses, and reviews, pre-registration and open-data practices) aimed at methodological rigour that is being encouraged in the scientific community.

Despite these undisputably positive efforts, we believe their potential to substantially improve the state of the art is limited without tackling the less "technical" influences hindering our understanding of phenomena of interest, and we are not alone in this opinion (Yarkoni, 2022; Scheel, 2022, Huber et al., 2023).

[17] Meehl considered experimenter error to make good theories look bad. We can reasonably conceive an opposite case. In addition, he believed the questionable validity of the instrument can go either way.

Meehl listed several influences that increase the logical distance between statistical results and claims they are interpreted to (dis)prove:

(1) loose derivation chain from theoretical premises to empirical observations providing an opportunity for logical slippage;

(2) problematic and taken-for-granted auxiliary theories i.e. explicit statements about the relationship between the theory and empirical facts;

(3) problematic ceteris paribus clause which refers to the unstated and untested expectation that randomization will take care of all of the hypothetical causal influences that vary in subjects so as they don't work in the opposite direction to the causal influence predicted by the theory and the auxiliary theories, and

(4) the crud factor refers to the "fact of the universe" in social sciences that everything correlates to some extent to everything else – the phenomena measured are multicausal and complex and most often have multiple explanations (Meehl, 1990).

The last fact, along with loose derivation chain, problematic auxiliary theories, and ceteris paribus clause, makes it very easy to explain the facts post hoc and fit them into almost any theory if we consider that a single finding based on null-hypothesis refutation confirms or refutes a theory.

Recently, the term "generalisability crisis" has been used to describe the state of treating "downstream symptoms" by big replication efforts and increased technical and statistical sophistication, instead of the logically preceding issue of generalisability (Yarkoni, 2022). Similarly to Meehl, Yarkoni (2022) elaborates on the consequences of a faulty logic of equating statistical models and verbal statements about psychological constructs and on the impracticality and conceptual impossibility of modeling everything as random effects, i.e. ensuring that the specific results are generalisable across the universe of hypothetical influences (tasks, subjects, individual differences, contextual differences). Although psychological constructs are verbal statements that are only as valid as their operationalization, it is a common practice to tacitly imply and behave as though a narrow operationalization is an acceptable proxy of an entire construct of interest. By equating a significant result of a narrowly operationalised construct with a verbal statement about a phenomenon (especially a contextually sensitive one) we are not describing the factual world but making *"sweeping generalisations (that) typically obtain little support from the reported empirical studies."* (Yarkoni, 2022, p.5).

All of the influences Meehl listed, underlying a widespread generalisability crisis in psychological science, can be readily identified in intergroup empathy bias research.

First, we undertook the task of systematising intergroup empathy bias research up to date, with a special focus on empathy-eliciting tasks and measurement and criteria for group differentiation. An attempt to make a comprehensive database of published registered reports in developmental research with a focus on the assessment of the hypotheses and their relationship to the results ended up abandoned after the pilot phase, as the author found most scientific claims throughout the papers so ill-defined that they cannot even be evaluated empirically – *"for most articles in the pilot sample, we were unable to establish exactly which test results informed which hypotheses and how they affected the authors' conclusions."* (Scheel, 2022, p. 2). We had an ambition of similar scope, but scaled it down to analysing how the authors defined empathy explicitly and what explanation they invoked to expect intergroup bias; otherwise, we would have abandoned our project as well.

Although it turned out to be manageable at the expense of depth, the task of analysing how empathy and bias were elicited and assessed proved to be unexpectedly daunting and resulted in cross-category counts and summaries of positive and negative findings that cannot be reliably attributed to any of the influences of interest, as studies that register bias and that do not vary in multiple dimensions we analysed and probably in many more we did not.

The explicit definitions of empathy were considerably general, mostly acknowledging the dual nature of empathy (affective vs. cognitive) or stressing its affective nature. Inclusion/exclusion criteria for empathic phenomena are implicit in all papers – therefore, the expected generalisability is also unuttered. Moreover, why a specific narrow phenomenon such as e.g. emotion identification or empathic concern was chosen to represent empathy defined broadly was seldom explained or elaborated on, but given as such. In other words, auxiliary theories about operationalizations are underspecified.

In "soft" psychology, every choice of an instrument or a task can be considered an auxiliary theory that the instrument/task is a good indicator of the trait/process intended to measure, and those can be as problematic as the main theory. In empathy research, there are several levels of those problems. Firstly, as there is no agreement even about the phenomena we label empathy, every operational definition of empathy is questionable if claimed to reflect the phenomenon in its entirety. However, that is frequently done, as illustrated in several conclusions and implications presented in the review study; in those not only the entire phenomenon is implicated, but its real-world consequences and application as well. Secondly, within a specific operationalization, the processes each task and measure invoke can depend on their specific combination and the external context. However, this issue is never tackled. Finally, each choice of measure includes tacit assumptions about the nature of empathy, depending on the author.

We will illustrate the loose derivation chain, problematic auxiliary theories, and ceteris paribus clause on the example of pain intensity and self-unpleasantness ratings.

Pain intensity assessment as an indicator of empathy was sampled from neuroscience studies that we partially replicated. However, both behavioural and neuroscience studies of physical pain almost universally use pain intensity assessment as an indicator of empathy. We have mentioned that it is not uncontested that emotion recognition should be considered empathy or its precondition. Factually, a pain intensity behavioural rating in these studies represents a conscious assessment of a sensory and emotional experience presented in the stimulus. An assumption most often implicit in the use of pain intensity ratings in empathy bias research is that the degree to which we share other's states will influence the amount of pain we attribute to a target. We argued that this assumption came from neuroscience studies that are specifically focused on affective sharing and operationally defined empathy as neural activation while disregarding behavioural responses. However, we can think of a number of processes other than affective sharing that contribute to participants' behavioural responses in each experiment. Moreover, these processes would differ from experiment to experiment, and increase in number as the stimulus complexity increases. It is mistaken to treat these ratings as equivalent. Let us illustrate with our studies why that is the case.

For example, in the experiment that compared painful vs. neutral faces, we were measuring emotion intensity identification. Painful and neutral events on neutral faces however measured emotional intensity atribution from symbolic signals of incoming pain. In Experiment 3 pain intensity ratings for painful accidents represented a complex assessment of the imagined painfulness of separate identity-unrelated events; in Experiment 4 pain intensity ratings for the painful fouls were complex assessments of visually presented identity-related painful events. The latter two experiments, therefore, differed in important contextual determinants, which exponentially increased the number of processes that could be influenced by group identity in shaping behavioural bias. In other words, even though we used the same label, each of the pain intensity ratings actually captured different processes in different tasks.

Similar arguments apply to self-unpleasantness ratings – the implicit hypothesis is that they depend on the degree of affective sharing. While that is certainly the case, they could also depend on myriad influences in different tasks. For example, it is possible that observing a painful face causes self-unpleasant feelings, if we assume that people automatically resonate with other's faces. However, self-unpleasantness ratings could also rely on e.g. empathic norms. It is less probable

however that any kind of empathic resonance influenced self-unpleasantness ratings in Experiment 2 where participants observed painful events happening on neutral faces, i.e. when the bodily signal of pain was absent. In this experiment, self-unpleasantness ratings could be alternatively influenced to fear of needles or more broadly, neuroticism. For self-unpleasantness is also important to note that while in neuroscience studies its use as an indicator of empathic reaction is not contested, some authors (e.g. Preston & DeWaal, 2002) explicitly exclude the responses that do not fulfil the criteria of self-other distinction from the term empathy.

We conducted a series of experiments to tackle the question of ingroup favouritism versus outgroup derogation as well as the ecological validity and contextual embeddedness of the empathy-eliciting stimuli in studying bias. The first two experiments represented a conceptual replication of two empathy-eliciting paradigms used in neuroscience studies – we labelled it conceptual as we had specifically chosen the groups and manipulated the stimuli to exclude social desirability and perceptual experience as explanations, and because we included a neutral group. In essence, however, we used identical or substantially similar testing paradigms. As previous studies are inconclusive, by addressing their limitations we aimed to provide benchmark expectations for these types of testing paradigms for intergroup bias. In the second two experiments, the pain was contextually embedded and either unrelated to group identity or directly related to "us and them" division – if there is an intergroup bias in empathic responses, we expected it to be more reliably manifested in contextualised form, especially if the context is identity-relevant.

In Experiment 3, we argued that the empathy-eliciting procedure is conceptually similar to misfortunes studies despite the differences in presentation modality. In both Experiment 3 and Experiment 4 we were governed by the contextual appropriateness in phrasing the questions. As previous studies of empathic responses to pain almost exclusively assessed pain intensity, other measures exploratory included – we wanted to test how they behaved and how they were mutually related. We labelled the congruent affective response empathic concern for communicative purposes (as it resembled the self-report questions from the Interpersonal Reactivity Index, but used congruent affective response alternately). However, one could argue for other labels, for example, compassion. In other words, in the absence of an etalon of empathy measurement, we, like the vast majority of other authors, chose the phrasing that we believe to represent a congruent affective response of the situation based on our implicit theory of what that response is. A separate issue in measuring higher-level empathic responses (in addition to not being certain which one is in and which one is out) is that we can differentiate them theoretically, but how appropriately we can transfer them into questions that participants can reliably respond in a differentiated manner? We do not dare to delve any deeper into this.

Results indicated that the issue of ingroup favouritism versus outgroup derogation got its share of "mixed" findings, precisely because there are different indicators of empathy (or in Meehl's terms different auxiliary theories). We interpreted the pattern of the results in terms of two social identity theory-derived accounts on bias: a cognitive one explaining the salience of group identity in our tasks in terms of its accessibility and fit, and a motivational one crossing the categories of identification and threat to explain why bias was observed only in Experiment 4.

We applied these frameworks post hoc, however. We started with social identity theory and its general premises about the strength of identification, ingroup love versus outgroup hate as distinguishable forms of bias, as well as the importance of social contextual variables for bias. As we have pointed out several times, both cognitive and motivational frameworks could incorporate several different patterns of findings, as they are, as many in psychological science, critically underspecified (Scheel, 2022). This is not to downplay the effort of the authors – on the contrary – both frameworks represent a theoretically sound explanation of when and how will social identity influence the way we see and interact with the social world and useful scaffolds to organise a body of findings on bias and coherently interpret them. In addition, "weak" theories explaining contextually determined phenomena cannot reasonably be expected to be directly related to

observable facts. But the point when they become useful is when they expand and elaborate on their key terms, and develop auxiliary theories (and not just summaries of research) about the expected structure of relationships to the observable phenomena, allowing the researchers to make more precise predictions. The other side of the coin is the perceivable lack of interest of the scientific community in theory development and theoretical and conceptual investigation and systematization of the empirical corpus, as well as for the systematic and thorough theoretical design of empirical studies.

Both Meehl (1990) and Yarkoni (2022) offer several suggestions on how the state can be improved, mainly revolving around increasing the understanding of and motivation for theoretical investigation and theoretical specification, as well as increasing the conceptual and methodological rigour and investing in high-quality complex data while refraining from overstated claims. Evidently, we subscribe to these analyses almost verbatim (at least the portion of them we possess the depth and breadth of knowledge to perform ourselves), even to the idea that purely destructive criticism can be beneficial. However, what we do not endorse is directing these recommendations primarily to individual persons, by listing the things we could do as individuals aspiring to be good scientists. Although both authors recognise the systemic problems of scientific incentives, the focus of their recommendations is aimed at what an individual should do.

Appeals to scientific integrity and campaigns of raising awareness about important issues may have value in shaping the behaviour of individual scientists – we do believe that most of us are (at least initially) primarily motivated by a desire to know. However, individuals (especially the ones in positions of power sympathetic to the arguments and concerns laid here) must in parallel aim to introduce systemic changes in publishing practices aimed at increasing the theoretical and conceptual rationale of empirical papers and incentivizing theoretical studies. These changes however must not remain individual initiatives incentivised only by good-scientist badges but should be lifted to the level of requirements. Although we strongly value the idea of a scientist as a being impervious to the motives of ordinary people and led only by scientific integrity and scientific reasoning in pursuit of truth, we are painfully aware that in present-day scientific research represents a loosely regulated domain of human activity (and employment) better characterised as governed by the laws of the free market. We believe the potential of individual behavioural changes to be limited absent the systemic efforts to incentivise theoretical research along with systematization and development of empirical concepts. More than the goodwill of a number of good researchers is needed to reshape the epistemic triangle in psychology to be more equilateral. Theoretical research is hard, and if the effort-to-reward ratio in terms of career benefits remains low, psychology is in danger of remaining a collection of loosely defined islets of empirical corpus dressed up in inferential statistics for the scientific rite. We must agree on the meaning of words first.

Empathy is a prime example of why it is important to define concepts precisely. It is a term that has spilled over from scientific to public discourse and has been additionally charged with value judgments – "sweeping generalisations" are descriptive not only of the way scientists discuss empathy studies but also of their real-world implications. Calls have been made to disregard empathy in moral decision-making (Bloom, 2016; Decety, 2021) due to an increasing number of studies revealing its parochial quality. However, before throwing the baby out with the bathwater, we believe empathy as a concept deserves a fair trial, impossible without proper systematization.

That is why we strongly contest the net summaries of vastly divergent studies on intergroup empathy bias. To borrow from the influential critique of intergroup bias in emotional face processing: *"Indeed, it is an attractive view for those who wish to propound cultural differences, and it has a catchy title. Such a view of the judgment process that is clearly not supported by data can easily lead to theories that advocate "fundamental" intergroup differences which, in turn, can easily polarize cultures against each other and contribute to the academic construction of walls and barriers among people that may be unnecessary and in any case unjustified by the literature.*

*Clearly, such developments have serious consequences for intergroup and interpersonal relations as well."* (Matsumoto, 2002, p. 242). In other words, claims labelled as scientific carry disproportional weight; it is our strong conviction they should not be made without appropriate evidence.

## Strengths, Limitations, and Recommendations

The main strength of the review study, in our view, is its focus on empathy measurement and conceptual analysis, i.e. on the relationship between empathy definition, operationalization, and interpretation. As argued in the previous chapter, the vagueness of concepts hinders progress in social sciences and conceptual analyses are sparse. We believe that this type of analytical effort should be beneficial to the field The study resulted in a comprehensive overview of empirical research that can serve as a starting point for further reviews, and that can be improved and updated in the future.

Limitations of the review study stem primarily from the lack of standardised guidelines or frameworks of analysis, in parallel with the sheer difficulty of the task (as we have already illustrated, for concepts like empathy it was a daunting task).

We aimed a) to establish if there is an effect of intergroup empathy bias as a differential empathic response to ingroups and outgroups and to b) analyze how authors conceive, measure, and interpret intergroup empathy bias. The former goal is similar to a typical review, and it implied that we had to limit our analysis to articles where baseline differences in empathy were reported. This meant omitting studies that did not report baseline differences in empathy, but that could legitimately be analysed conceptually. In addition, considering the scope of the term empathy, we decided to select only those articles where the authors explicitly label what they measure as an empathic response, thereby excluding studies where authors did not use the term but essentially measured a similar response (emotional face perception studies are a good example). Finally, we also touched upon the scope of the hypothesised implications of empathy studies and the manner they were typically communicated, which is a topic that deserves its own review paper.

In hindsight, it might have been more informative if we had focused on a single goal – we believe the end result might have been a more comprehensive study. This is our main recommendation for further reviews. In addition, we recommend expanding the analysis to all studies of empathy-related phenomena regardless of the authors' use of the label, to gain further insight into the social contextual sensitivity of specific empathic processes. If integrated with theoretical accounts of empathy and social biases, these conceptual reviews have the potential to improve our knowledge of both. Finally, we recommend introducing and maintaining systemic efforts to incentivise and improve conceptual analyses, especially if the topic is a contextually defined phenomenon such as intergroup empathy bias. We further encourage structuring and publicly sharing the conceptual review materials and their further expansion, which is also our goal for the future.

We will highlight several theoretical and methodological strengths of the empirical studies described in this thesis.

Firstly, with our choice of group identity and careful design of the stimuli, we tackled the issue of social desirability in intergroup empathy bias studies, i.e. the design was such that it excluded this alternative explanation.

Secondly, the sample size in our experiments represents an improvement compared to most comparable previous studies. This goes to sample composition as well, albeit to a lesser degree.

Finally and most importantly, the main strength of our studies is their individual and joint theoretical embeddedness and interpretation. Experiments 1 and 2 were designed to contribute to the theoretical discussion about the relationship between behavioural and neural indicators. Their results further point out that common social desirability explanations for the absence of intergroup empathy bias in behavioural indicators in neuroscience studies do not stand. In Experiments 3 and 4 the scope of assessed empathic reactions was broadened and the empathy-eliciting painful events were contextualised in a theoretically relevant manner.

The absence of bias in Experiments 1-3 and its strong presence in Experiment 4 implicated that when investigating social phenomena one needs to take into account the nature of the task presented to participants, i.e. to theoretically and conceptually analyse and not tacitly imply its social nature. In addition, our joint interpretation of results in terms of cognitive and motivational frameworks developed within the social identity theory of bias as well as in terms of general empathy frameworks highlighted the contextualised and motivated nature of social biases as such and pointed to several gaps in theory development that needed addressing. Lastly, we reflected on and identified how conceptual disorder in empathy research manifested in intergroup empathy bias studies and explicitly pointed out several concepts and related methodological choices taken for granted that need explicit analysis and discussion.

Limitations of our studies at the same time represent recommendations for further empirical studies of empathic responses to physically painful stimuli.

Experiments 1-4 were interpreted theoretically but not designed in advance to test specific predictions of social identity frameworks about empathy for pain bias. In parallel with theoretical development enabling those predictions, future studies should be designed to test them.

Experiments 1 and 2 were direct replications of the most frequent empathy-eliciting paradigms in neuroscience. Future studies of this kind could include a broader scope of empathic responses (e.g. congruent affect or empathic concern), as well as more nuanced measurement techniques than Likert scale ratings (e.g., visual sliders, response times, etc.) (preceded by conceptual consideration and further empirical investigation of appropriate responses and measures). In addition, decontextualised tasks such as these would benefit from open-ended feedback about participants' response strategies. Finally, it would be theoretically valuable to investigate if presenting the empathy-eliciting tasks as identity-relevant by changing the instructions would lead to the emergence of intergroup bias and how different kinds of responses would be affected.

Experiment 3 was structurally most similar to a comparatively large group of studies measuring congruent and incongruent affective responses (*how good/bad does that make you feel?*) to various misfortunes of others, and those studies observed universal intergroup empathy bias. Our study, focusing exclusively on physically painful misfortunes, did not. Compared to misfortunes studies that mainly used visual sliders, we used Likert scale ratings in all studies. Future studies should consider including more sensitive measures to confirm or dispute our findings.

In Experiment 4 we used images of real-life events as empathy-eliciting stimuli. While highlighting ecological validity and contextual embeddedness, we sacrificed the unambiguousness of stimuli to a certain extent. As we argued this choice to be theoretically valuable for studying social biases, our recommendation for further studies is to continue with this tradeoff (after more extensive pilot testing of the participant's perception of the stimuli) and to make their complex stimuli and appended ratings publicly available.

## Concluding Remarks

To understand the meaning of the concept of intergroup empathy bias with respect to the definition, measurement, and interpretation of empathy, we conducted a review study. We argued that the results spoke in favour of the idea that the notion originated primarily from neuroscientific studies of empathy for physical pain and that these studies still represent the most important part of the empirical corpus in studying intergroup empathy bias, thereby confirming our preliminary expectations. Moreover, we pointed out the heterogeneity in theoretical and operational definitions of empathy in intergroup empathy bias research and criticised the ubiquitous praxis of implicitly and explicitly generalising the results. In our view, it hinders not only the clarity of the concept in the literature but the real-world implications of this research as well. The most important conclusion of the review study is not about the specifics of intergroup empathy bias but pertains to a general need for conceptual development and process specification of empathic phenomena, as well as for relying on general socio-psychological theories of bias in both design and interpretation in a more deliberate and explicit fashion.

Empirical studies were designed to tackle some of the issues we identified in the review. Specifically, in two experiments we conceptually replicated frequently used empathy-eliciting paradigms in neuroscience, tackling the alternative social desirability explanation for the absence of bias in behavioural responses, as well as small sample sizes. In another two experiments, we investigated the ecological validity and contextual embeddedness of the empathy-eliciting event. Our main conclusion is that biased empathic responses are contextually defined phenomena and that intergroup empathy bias will reliably emerge only when the group identity is highly salient and the biased response is functional for navigating the social world. The central contribution of this thesis is that when studying social phenomena, we must take care about what exactly we are asking our participants. Theoretical interpretation of the series of studies further highlighted the need to study social phenomena in context as well as the need to develop better and more precise theoretical expectations about both empathy and bias.

We believe that conceptual investigations aiming to build consensus on theoretical and operational definitions of psychological phenomena are of crucial importance for advancing scientific knowledge. For the concept of empathy, they are a necessity. Presently, empathy is everything, anything, and nothing at all, a pseudo-scientific concept and a buzzword uncritically employed in public discourse. However, the questions we are trying to address by studying empathic responses are too important for conceptual issues to be ignored. How do we get to know the internal states of others and how do we use that knowledge to navigate the social world are fundamental questions of human social life. They deserve much better answers than we can presently offer. To move forward, we need to take a step back and build a conceptual foundation. Only then can scientific knowledge be used to make the world a (bit) nicer place.

## Literature

References that were included in the review study are marked with R.


Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience, 22*(1), 97–108. https://doi.org/10.1162/jocn.2009.21187

Abu-Akel, A., Fischer-Shofty, M., Levkovitz, Y., Decety, J., & Shamay-Tsoory, S. (2014). The role of oxytocin in empathy to the pain of conflictual out-group members among patients with schizophrenia. *Psychological medicine, 44*(16), 3523-3532. https://doi.org/10.1017/S003329171400097X R

Ainslie, G. & Monterosso, J. (2002). Hyperbolic discounting lets empathy be a motivated process. [Peer commentary on "Empathy: Its ultimate and proximate bases" by S. Preston & F. De Waal]. *Behavioral and Brain Sciences, 25*, 20-21. http://dx.doi.org/10.1017/S0140525X02000018

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience, 15*(10), 670–682. https://doi.org/10.1038/nrn3800

Andersson, H. E. (2010). What Activates an Identity? The Case of Norden. *International Relations*, *24*(1), 46-64. https://doi.org/10.1177/0047117809359039

Arroyo, J. A. (1996). Psychotherapist bias with Hispanics: An analog study. *Hispanic Journal of Behavioral Sciences, 18*(1), 21-28. https://doi.org/10.1177/07399863960181003 R

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, *20*(11), 1018-1022. https://doi.org/10.1016/j.cub.2010.03.071 R

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 R

Bandura, A. (2002). Reflexive empathy: On predicting more than has ever been observed. [Peer commentary on "Empathy: Its ultimate and proximate bases" by S. Preston & F. De Waal]. *Behavioral and Brain Sciences, 25*, 24-25. http://dx.doi.org/10.1017/S0140525X02000018

Bartholow, B. D., & Ito, T. A. (2009). The Neural Correlates of Race. *Trends in Cognitive Sciences, 13*(12), 524–531. https://doi.org/10.1016/j.tics.2009.10.002

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *Social neuroscience. The social neuroscience of empathy* (p. 3–15). MIT Press.

Berlingeri, M., Gallucci, M., Danelli, L., Forgiarini, M., Sberna, M., & Paulesu, E. (2016). Guess who's coming to dinner: brain signatures of racially biased and politically correct behaviors. *Neuroscience, 332,* 231-241. https://doi.org/10.1016/j.neuroscience.2016.06.048 R

Bernache-Assollant, I., Bouchet, P., & Lacassagne, M.-F. (2007). Spectators' Identification with French Sport Teams; A French Adaptation of the Sport Spectator Identification Scale. *Perceptual and Motor Skills, 104*(1), 83–90. https://doi.org/10.2466/pms.104.1.83-90

Berry, D. R. (2017). Bridging the empathy gap: Effects of brief mindfulness training on helping outgroup members in need. [Unpublished Doctoral dissertation]. Virginia Commonwealth University. https://doi.org/10.25772/HGBR-6504 R

Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: Multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers Robotics and AI, 5*. https://doi.org/10.3389/frobt.2018.00026

Blanca, M. J., Arnau Gras, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: impact on type I error and power. *Psicothema, 35*(1), 21-29. https://doi.org/10.7334/psicothema2022.292

Bloom, P. (2016). *Against Empathy: The Case for Rational Compassion*. HarperCollins Publishers.

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

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 R

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 R

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

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376. https://doi.org/10.1038/nrn3475

Cameron, C. D. (2018). Motivating empathy: Three methodological recommendations for mapping empathy. *Social and Personality Psychology Compass*, *12*(11), e12418. https://doi.org/10.1111/spc3.12418

Canevello, A., Hall, J., & Walsh, J. I. (2022). Empathy-mediated altruism in intergroup contexts: The roles of posttraumatic stress and posttraumatic growth. *Emotion, 22*(8), 1699–1712. https://doi.org/10.1037/emo0000803 R

Cao, Y., Contreras-Huerta, L. S., McFadyen, J., & Cunnington, R. (2015). Racial bias in neural response to others' pain is reduced with other-race contact. *Cortex, 70*, 68-78. https://doi.org/10.1016/j.cortex.2015.02.010 R

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior research methods*, *44*, 1255-1265. https://doi.org/10.3758/s13428-012-0186-0

Carp, J. (2012a). On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. *Frontiers in neuroscience*, *6*, 149. https://doi.org/10.3389/fnins.2012.00149

Carp, J. (2012b). The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, *63*(1), 289-300. https://doi.org/10.1016/j.neuroimage.2012.07.004

Chen, C., Crivelli, C., Garrod, O., Fernandez-Dols, J. M., Schyns, P., & Jack, R. (2016). Facial expressions of pain and pleasure are highly distinct. *Journal of Vision*, *16*(12), 210.

Chen, J., Wu, K., Shi, Y., & Ai, X. (2021). The relationship between dispositional self-construal and empathy for ingroup and outgroup members' pain: evidence from ERPs. *Acta Psychologica Sinica*, *53*(6), 629. https://doi.org/10.3724/SP.J.1041.2021.00629 R

Cheon, B. K., Im, D. M., 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 R

Christopher, J. C., Wendt, D. C., Marecek, J., & Goodman, D. M. (2014). Critical cultural awareness: Contributions to a globalizing psychology. *American Psychologist*, *69*(7), 645-655. https://doi.org/10.1037/a0036851

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. (2018). Pleasure in response to out-group pain as a motivator of intergroup aggression. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology*. New York, USA: Guilford Press.

Cikara, M., & Fiske, S. T. (2011). Bounded Empathy: Neural Responses to Outgroup Targets' (Mis)fortunes. *Journal of Cognitive Neuroscience*, *23*(12), 3791–3803. https://doi.org/10.1162/jocn_a_00069

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

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus Them: Social Identity Shapes Neural Responses to Intergroup Competition and Harm. *Psychological Science*, *22*(3), 306–313. https://doi.org/10.1177/0956797610397667

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 R

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 & Biobehavioral Reviews*, *83*, 132-139. https://doi.org/10.1016/j.neubiorev.2017.10.009

Coll, M. P. (2018). Meta-analysis of ERP investigations of pain empathy underlines methodological issues in ERP research. *Social cognitive and affective neuroscience*, *13*(10), 1003-1017. https://doi.org/10.1093/scan/nsy072

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

Contreras-Huerta, L. S., Hielscher, E., Sherwell, C. S., Rens, N., & Cunnington, R. (2014). Intergroup relationships do not reduce racial bias in empathic neural responses to pain. *Neuropsychologia*, *64*, 263-270. https://doi.org/10.1016/j.neuropsychologia.2014.09.045 R

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2014). Empathy: A Review of the Concept. *Emotion Review, 8*(2), 144–153. https://doi.org/10.1177/1754073914558466

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

De Dreu, C. K. W. (2010). Social conflict: The emergence and consequences of struggle and negotiation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (p. 983–1023). John Wiley & Sons. https://doi.org/10.1002/9780470561119.socpsy002027

De Jong, D., Hortensius, R., Hsieh, T. Y., & Cross, E. S. (2021). Empathy and schadenfreude in human–robot teams. Journal of Cognition, 4(1), https://doi.org/10.5334/joc.177 R

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science, 9*(4), 525–537. https://doi.org/10.1177/1745691614545130

Dildine, T. C., & Atlas, L. Y. (2019). The need for diversity in research on facial expressions of pain. *Pain*, *160*(8), 1901. https://doi.org/10.1097/j.pain.0000000000001593

Dovidio, J. F., Johnson, J. D., Gaertner, S. L., Pearson, A. R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and Intergroup Relations. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 393–408). American Psychological Association.

Drimalla, H., Landwehr, N., Hess, U., & Dziobek, I. (2019). From face to face: the contribution of facial mimicry to cognitive and emotional empathy. *Cognition and Emotion*, *33*(8), 1672-1686. https://doi.org/10.1080/02699931.2019.1596068

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 R

Eisenberg, N., Eggum, N. D., & Di Giunta, L. (2010). Empathy-related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, *4*(1), 143–180. https://doi.org/10.1111/j.1751-2409.2010.01020.x

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712–717. https://doi.org/10.1037/0022-3514.53.4.712

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual review of psychology*, *53*(1), 161-186. https://doi.org/10.1146/annurev.psych.53.100901.135228

Eres, R., & Molenberghs, P. (2013). The influence of group membership on the neural correlates involved in empathy. *Frontiers in Human Neuroscience*, *7*, 176. https://doi.org/10.3389/fnhum.2013.00176

Fabi, S., & Leuthold, H. (2018). Racial bias in empathy: Do we process dark-and fair-colored hands in pain differently? An EEG study. *Neuropsychologia*, *114*, 143-157. https://doi.org/10.1016/j.neuropsychologia.2018.04.024 R

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.4324/9781315187280-7

Forgiarini, M., Gallucci, M., & Maravita, A. (2011). Racism and the empathy for pain on our skin. *Frontiers in Psychology*, *2*, 108. https://doi.org/10.3389/fpsyg.2011.00108 R

Fourie, M. M., Stein, D. J., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2017). Empathy and moral emotions in post-apartheid South Africa: an fMRI investigation. *Social Cognitive and Affective Neuroscience*, *12*(6), 881-892. https://doi.org/10.1093/scan/nsx019 R

Fourie, M. M., Stein, D. J., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2019). Effects of early adversity and social discrimination on empathy for complex mental states: An fMRI investigation. *Scientific reports*, *9*(1), 12959. https://doi.org/10.1038/s41598-019-49298-4 R

Fourie, M. M., Subramoney, S., & Gobodo-Madikizela. (2017). A Less Attractive Feature of Empathy: Intergroup Empathy Bias. In M. Kondo (Ed.), *Empathy: An Evidence-based Interdisciplinary Perspective* (pp. 45–62). Rijeka, Croatia: InTech.

Gamsakhurdashvili, D., Antov, M. I., & Stockhorst, U. (2021). Sex-hormone status and emotional processing in healthy women. *Psychoneuroendocrinology, 130*, 105258. https://doi.org/10.1016/j.psyneuen.2021.105258 R

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Grimm, F. E. (2016). Predicting and Intervening Intergroup Empathy Bias in Soccer Fans. [Unpublished master's thesis]. Universität Wien, https://doi.org/10.25365/thesis.44635 R

Guthridge, M., & Giummarra, M. J. (2021). The taxonomy of empathy: A meta-definition and the nine dimensions of the empathic system. *Journal of Humanistic Psychology*, 00221678211018015. https://doi.org/10.1177/00221678211018015

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 R

Gutsell, J. N., Simon, J. C., & Jiang, Y. (2020). Perspective taking reduces group biases in sensorimotor resonance. *Cortex, 131,* 42-53. https://doi.org/10.1016/j.cortex.2020.04.037 R

Hall, J., & Schwartz, R. (2018). Empathy present and future. *The Journal of Social Psychology, 159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442

Han, S. (2018). Neurocognitive Basis of Racial Ingroup Bias in Empathy. *Trends in Cognitive Sciences, 22*(5), 400–421. https://doi.org/0.1016/j.tics.2018.02.013

Han, X., Luo, S., & Han, S. (2015). Embodied neural responses to others' suffering. *Cognitive Neuroscience, 7*(1–4), 114–127. https://doi.org/10.1080/17588928.2015.1053440

Hanson, E. (2017). On the Ironic Effects of Being Empathic: Consequences for Attitude Polarization and Intergroup Conflict. [Unpublished master's thesis]. Washington University in St. Louis, https://openscholarship.wustl.edu/art_sci_etds/1063 R

Haslam, S. A., Oakes, P. J., McGarty, C., Turner, J. C., & Onorato, R. S. (1995). Contextual changes in the prototypicality of extreme and moderate outgroup members. *European Journal of Social Psychology, 25*(5), 509-530. https://doi.org/10.1002/ejsp.2420250504

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 R

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

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology, 58A*(2), 193–233. https://doi.org/10.1080/02724980443000502

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

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., ... & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO$_7$ scale. *Journal of personality and social psychology*, *109*(6), 1003. https://doi.org/10.1037/pspi0000033

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences*, *113*(16), 4296-4301. https://doi.org/10.1073/pnas.1516047113

Hornsey, M. J. (2008). Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass, 2*, pp. 204–222. https://doi.org/10.1111/j.1751-9004.2007.00066.x

Huang, S., & Han, S. (2014). Shared beliefs enhance shared feelings: religious/irreligious identifications modulate empathic neural responses. *Social Neuroscience*, *9*(6), 639-649. https://doi.org/10.1080/17470919.2014.934396 R

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., ... & Holzmeister, F. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences*, *120*(23), e2215572120. https://doi.org/10.1073/pnas.2215572120

Jackson, J. W. (1993). Realistic group conflict theory: A review and evaluation of the theoretical and empirical literature. *The Psychological Record*, *43*(3), 395-414.

Jiang, C., Varnum, M. E. W., Hou, Y., & Han, S. (2014). Distinct effects of self-construal priming on empathic neural responses in Chinese and Westerners. *Social Neuroscience, 9*(2), 130–138. https://doi.org/10.1146/annurev.psych.53.100901.135109

Johnson, S. L., Hritz, A. C., Royer, C. E., & Blume, J. H. (2016). When Empathy Bites Back: Cautionary Tales from Neuroscience for Capital Sentencing. *Fordham Law Review*, *85*(2), 573–598.

Karakaš, T. (2020), Utjecaj etničke pripadnosti aktera u scenariju na samoprocjenjenu empatiju. [Unpublished master's thesis]. Josip Juraj Strossmayer University of Osijek. https://urn.nsk.hr/urn:nbn:hr:142:962065 R

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492–2502. https://doi.org/10.1016/j.neuroimage.2010.10.014

Leiner, D. J. (2019, December). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229-248. https://doi.org/10.18148/srm/2019.v13i3.7403

Levin, S., Federico, C. M., Sidanius, J., & Rabinowitz, J. L. (2002). Social dominance orientation and intergroup bias: The legitimation of favoritism for high-status groups. *Personality and Social Psychology Bulletin, 28*(2), 144-157. https://doi.org/10.1177/0146167202282002

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology, 49*(4), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013

Li, X., Liu, Y., Luo, S., Wu, B., Wu, X., & Han, S. (2015). Mortality salience enhances racial in-group bias in empathic neural responses to others' suffering. *NeuroImage, 118*, 376-385. https://doi.org/10.1016/j.neuroimage.2015.06.023 R

Luo, S., Han, X., Du, N., & Han, S. (2018). Physical coldness enhances racial in-group bias in empathy: Electrophysiological evidence. *Neuropsychologia, 116*, 117-125. https://doi.org/10.1016/j.neuropsychologia.2017.05.002 R

Luo, S., Li, B., Ma, Y., Zhang, W., Rao, Y., & Han, S. (2015). Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. *NeuroImage, 110*, 22-31. https://doi.org/10.1016/j.neuroimage.2015.01.042 R

Luo, S., Zhang, T., Li, W., Yu, M., Hein, G., & Han, S. (2019). Interactions between oxytocin receptor gene and intergroup relationship on empathic neural responses to others' pain. *Social Cognitive and Affective Neuroscience, 14*(5), 505-517. https://doi.org/10.1093/scan/nsz029 R

Machado, A., Lourenço, O., & Silva, F. J. (2000). Facts, concepts, and theories: The shape of psychology's epistemic triangle. *Behavior and Philosophy*, 1-40. https://www.jstor.org/stable/27759402

Mackie, D. M., Smith, E. R., & Ray, D. G. (2008). Intergroup Emotions and Intergroup Relations. *Social and Personality Psychology Compass, 2*(5), 1866–1880. https://doi.org/10.1111/j.1751-9004.2008.00130.x

Mathur, V. A., Harada, T., & Chiao, J. Y. (2011). Racial Identification Modulates Default Network Activity for Same and Other Races. *Human Brain Mapping, 33*(8), 1883–1893. https://doi.org/10.1002/hbm.21330

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 R

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of cross-cultural psychology, 23*(1), 72-84. https://doi.org/10.1177/0022022192231005

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: Comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin, 128*(2), 236–242. https://doi.org/10.1037/0033-2909.128.2.236

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and emotion*, *13*, 143-157. https://doi.org/10.1007/BF00992959

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological reports*, *66*(1), 195-244. https://doi.org/10.2466/pr0.1990.66.1.195

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharzad, A. (2020). The Delaware Pain Database: a set of painful expressions and corresponding norming data. *Pain reports*, *5*(6), e853. https://doi.org/10.1097/PR9.0000000000000853

Milivojević, T. Đ. (2015). Inflacija upotrebe reči empatija i emocionalizacija etike. *Komunikacija i Kultura Online*, *6*(6), 277–305.

Miller, E. T., & Abu-Alhaija, D. M. (2019). Cultural influences on pain perception and management. *Pain Management Nursing*, *20*(3), 183-184. https://doi.org/10.1016/j.pmn.2019.04.006

Miron-Shatz, T., Ormianer, M., Rabinowitz, J., Hanoch, Y., & Tsafrir, A. (2020). Physician experience is associated with greater underestimation of patient pain. *Patient education and counseling*, *103*(2), 405-409. https://doi.org/10.1016/j.pec.2019.08.040 R

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience and Biobehavioral Reviews, 37(*8), 1530–1536, https://doi.org/10.1016/j.neubiorev.2013.06.002

Molenberghs, P., & Morrison, S. (2014). The role of the medial prefrontal cortex in social categorization. *Social Cognitive and Affective Neuroscience*, *9*(3), 292–296. https://doi.org/10.1093/scan/nss135

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*, 389. https://doi.org/10.3389/fpsyg.2012.00389 R

Moroń, M., Biolik-Moroń, M., & Matuszewski, K. (2021). Alterations in Religious Rituals Due to COVID-19 Could Be Related to Intragroup Negativity: A Case of Changes in Receiving Holy Communion in the Roman Catholic Community in Poland. *Religions*, *12*(4), 240. https://doi.org/10.3390/rel12040240 R

Neumann, D. L., Boyle, G. J., & Chan, R. C. (2013). Empathy towards individuals of the same and different ethnicity when depicted in negative and positive contexts. *Personality and Individual Differences*, *55*(1), 8-13. https://doi.org/10.1016/j.paid.2013.01.022 R

Neumann, D. L., Chan, R. C. K., Boyle, G. J., Wang, Y., & Westbury, H. R. (2014). Measures of Empathy: Self-Report, Behavioral, and Neuroscientific Approaches. In G. Boyle, D. Saklofske, & G. Matthews (Eds.), *Measures of Personality and Social Psychological Constructs* (1st ed., pp. 257–289). London, UK: Academic Press.

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

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Ozkara, B. Y. (2021). The neural substrate of schadenfreude: The effects of competition level changes on the processing of pain in others. *New Ideas in Psychology, 62*, 100853. https://doi.org/10.1016/j.newideapsych.2021.100853 R

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. A., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting. P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery, 88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Pastoors, J. K. (2020). The Opposite of Love: How Parasocial Interaction with NPC's Can be Enhanced Through Intergroup Competition (Master thesis, Breda University of Applied Sciences). https://doi.org/10.13140/RG.2.2.31175.52642 R

Paterson, J., Brown, R., & Walters, M. (2018). Understanding victim group responses to hate crime: Shared identities, perceived similarity and intergroup emotions. *TPM. Testing, Psychometrics, Methodology in Applied Psychology, 25*(2), 163-177. https://doi.org/10.4473/TPM25.2.1 R

Paul, M., Govaart, G. H., & Schettino, A. (2021). Making ERP research more transparent: guidelines for preregistration. *International Journal of Psychophysiology, 164*, 52-63. https://doi.org/10.1016/j.ijpsycho.2021.02.016

Peacock, S., & Patel, S. (2008). Cultural influences on pain. *Reviews in pain, 1*(2), 6-9. https://doi.org/10.1177/204946370800100203

Petsnik, C., & Vorauer, J. D. (2020). Do dominant group members have different emotional responses to observing dominant-on-dominant versus dominant-on-disadvantaged ostracism? Some evidence for heightened reactivity to potentially discriminatory ingroup behavior. *PLoS ONE, 15*(6), e0234540. https://doi.org/10.1371/journal.pone.0234540 R

Phillips, C. J. (2021). Effects of Patient's Race on Pain Perception and Treatment in Nursing Students. [Unpublished Thesis]. University of Southern Mississippi, https://aquila.usm.edu/honors_theses/774 R

Pirandelo, L. (2007). *Jedan, nijedan i sto hiljada* (A. Levi & M. Radosavljević, Trans.). PAIDEIA. (Original work published 1926)

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology, 67*(4), 741. https://doi.org/10.1037/0022-3514.67.4.741

Preston, S. D., & De Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1–72. https://doi.org/10.1017/S0140525X02000018

Prkachin, K. M. (2009). Assessing pain by facial expression: facial expression as nexus. *Pain Research and Management, 14,* 53-58. https://doi.org/10.1155/2009/542964

Richins, M. T. (2017). Intergroup Empathy - Beyond Boundaries. [Unpublished Doctoral dissertation]. University of Exter. https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31148/RichinsM.pdf?sequence=3&isAllowed=y

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 R

Riečanský, I., Lengersdorff, L. L., Pfabigan, D. M., & Lamm, C. (2020). Increasing self-other bodily overlap increases sensorimotor resonance to others' pain. *Cognitive, Affective, & Behavioral Neuroscience, 20,* 19-33. https://doi.org/10.3758/s13415-019-00724-0 R

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 R

Ruckmann, J., Bodden, M., Jansen, A., Kircher, T., Dodel, R., & Rief, W. (2015). How pain empathy depends on ingroup/outgroup decisions: a functional magnet resonance imaging study. *Psychiatry Research: Neuroimaging, 234*(1), 57-65. https://doi.org/10.1016/j.pscychresns.2015.08.006 R

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development, 31*(1), e2295. https://doi.org/10.1002/icd.2295

Scheepers, D., & Derks, B. (2016). Revisiting social identity theory from a neuroscience perspective. *Current Opinion in Psychology, 11,* 74–78. https://doi.org/10.1016/j.copsyc.2016.06.006

Schimmack, U. (2021). The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science, 16*(2), 396-414. https://doi.org/10.1177/1745691619863798

Schimmack, U. (2021). Invalid claims about the validity of implicit association tests by prisoners of the implicit social-cognition paradigm. *Perspectives on Psychological Science, 16*(2), 435-442. https://doi.org/10.1177/1745691621991860

Schurtz, D.R., Combs, D., Hoogland, C., & Smith, R H. (2014). Schadenfreude in sports and politics: Social identity perspective. In: W. W. van Dijk & J. W. Ouwerkerk (Eds.), *Schadenfreude: Understanding pleasure at the misfortune of others* (pp. 170-185). Cambridge, UK: Cambridge University Press.

Sessa, P., Meconi, F., Castelli, L., & Dell'Aqua, R. (2014). Taking one's time in feeling other-race pain: an event-related potential investigation on the time-course of cross-racial empathy. *Social Cognitive and Affective Neuroscience, 9*(4), 454–463. https://doi.org/10.1093/scan/nst003

Shamay-Tsoory, S. G., Abu-Akel, A., Palgi, S., Sulieman, R., Fischer-Shofty, M., Levkovitz, Y., & Decety, J. (2013). Giving peace a chance: Oxytocin increases empathy to pain in the context of the Israeli–Palestinian conflict. *Psychoneuroendocrinology, 38*(12), 3139-3144. https://doi.org/10.1016/j.psyneuen.2013.09.015 R

Sharifian, M., Hatami, J., Batouli, S. A. H., & Boroujeni, M. M. F. (2022). Citizens of the world: National stereotypes do not affect empathic response in the presence of individuating information. *International Journal of Psychology, 57*(2), 251-260. https://doi.org/10.1002/ijop.12807 R

Shen, L. (2010). *State Empathy Scale* [Database record]. APA PsycTests. https://doi.org/10.1037/t81376-000

Shen, F., Hu, Y., Fan, M., Wang, H., & Wang, Z. (2018). Racial bias in neural response for pain is modulated by minimal group. *Frontiers in Human Neuroscience, 11*, 661. https://doi.org/10.3389/fnhum.2017.00661 R

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 R

Sheng, F., Du, N., & Han, S. (2017). Degraded perceptual and affective processing of racial out-groups: An electrophysiological approach. *Social neuroscience, 12*(4), 479-487. https://doi.org/10.1080/17470919.2016.1182944 R

Sheng, F., Han, X., & Han, S. (2015). Dissociated Neural Representations of Pain Expressions of Different Races. *Cerebral Cortex, 26*(3), 1221–1233. https://doi.org/10.1093/cercor/bhu314

Sheng, F., Liu, Q., Li, H., Fang, F., & Han, S. (2014). Task modulations of racial bias in neural responses to others' suffering. *NeuroImage, 88*, 263-270. https://doi.org/10.1016/j.neuroimage.2013.10.017 R

Sheng, F., Liu, Y., Zhou, B., Zhou, W., & Han, S. (2013). Oxytocin modulates the racial bias in neural responses to others' suffering. *Biological psychology, 92*(2), 380-386. https://doi.org/10.1016/j.biopsycho.2012.11.018 R

Singal, J. (2021). *The quick fix: Why fad psychology can't cure our social ills*. Farrar, Straus and Giroux.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*(1), 81-96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Smith, E. R. , & Mackie, D. M. (2008). Intergroup emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 428 – 439). New York, NY: Guilford.

Spinda, J. S. (2011). The development of basking in reflected glory (BIRGing) and cutting off reflected failure (CORFing) measures. *Journal of sport Behavior, 34*(4), 392-420.

Stürmer, S., Snyder, M., & Omoto, A. M. (2005). Prosocial Emotions and Helping: The Moderating Role of Group Membership. *Journal of Personality and Social Psychology, 88*(3), 532–546. https://doi.org/10.1037/0022-3514.88.3.532

Stürmer, S., Snyder, M., Kropp, A., & Siem, B. (2006). Empathy-Motivated Helping: The Moderating Role of Group Membership. *Personality and Social Psychology Bulletin, 32*, 943–956. https://doi.org/10.1177/0146167206287363

Suleiman, R., Yahya, R., Decety, J., & Shamay-Tsoory, S. (2018). The impact of implicitly and explicitly primed ingroup–outgroup categorization on the evaluation of others pain: The case of the Jewish–Arab conflict. *Motivation and Emotion, 42*, 438-445. https://doi.org/10.1007/s11031-018-9677-3 R

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual review of psychology*, *33*(1), 1-39.

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). Boston, MA: Brooks/Cole.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In J. T. Jost & J. Sidanius (Eds.), *Key readings in social psychology: Political psychology*. (pp. 276–293). https://doi.org/10.4324/9780203505984-16

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

Tingley, D. (2006). Neurological imaging as evidence in political science: A review, critique, and guiding assessment. *Social Science Information, 45*(1), 5–33. https://doi.org/10.1177/0539018406061100

Trawalter, S., Hoffman, K. M., & Waytz, A. (2012). Racial bias in perceptions of others' pain. PloS one, 7(11), e48546. https://doi.org/10.1371/journal.pone.0152334 R

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), 62. https://doi.org/10.1038/s42003-018-0073-z

Vaes, J., Meconi, F., Sessa, P., & Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia, 89*, 132-140. https://doi.org/10.1016/j.neuropsychologia.2016.06.004 R

Vaughn, D. A., Savjani, R. R., Cohen, M. S., & Eagleman, D. M. (2018). Empathic neural responses predict group allegiance. *Frontiers in Human Neuroscience, 12*, 302. https://doi.org/10.3389/fnhum.2018.00302 R

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 R

Wang, C., Wu, B., Liu, Y., Wu, X., & Han, S. (2015). Challenging emotional prejudice by changing self-concept: priming independent self-construal reduces racial in-group bias in neural responses to other's pain. *Social cognitive and affective neuroscience*, *10*(9), 1195-1201. https://doi.org/10.1093/scan/nsv005 R

Wann, D. L., & Branscombe, N. R. (1993). Sports fans: Measuring degree of identification with their team. *International journal of sport psychology*, *24*(1), 1–17.

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

Westbury, H. R., & Neumann, D. L. (2008). Empathy-related responses to moving film stimuli depicting human and non-human animal targets in negative circumstances. Biological psychology, 78(1), 66-74. https://doi.org/10.1016/j.biopsycho.2007.12.009 R

Wilkinson, D., & Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews Neuroscience*, *5*, 67–73. https://doi.org/10.1038/nrn1302

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 R

Xu, Y., Chen, S., Kong, Q., & Luo, S. (2021). The residential stability mindset increases racial in-group bias in empathy. *Biological Psychology*, *165*, 108194. https://doi.org/10.1016/j.biopsycho.2021.108194 R

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Zaki, J. (2014). Empathy: a motivated account. *Psychological bulletin*, *140*(6), 1608. https://doi.org/10.1037/a0037679

Zaki, J., & Cikara, M. (2015). Addressing Empathic Failures. *Current Directions in Psychological Science*, *24*(6), 471–476. https://doi.org/10.1177/0963721415599978

# Appendices

Appendix A

Search strategies per database

We conducted the literature search in three databases: ScienceDirect, Google Scholar, PsycInfo. However, the search forms in each of these databases are structurally different, which resulted in somewhat different search strategies.

ScienceDirect offers the possibility to limit your search to title, abstract and keywords, which we believe to be both specific enough to return mostly relevant articles and broad enough to guarantee most or all IEB studies will be included. As our goal is to understand how IEB is conceptualised in the current scientific practice, we believe it is justified limit the search only to the summary parts of the articles and screen only the records which contain the search terms in their title, abstract, or keywords. However, we allowed the search terms to appear separately from one another to ensure all of the relevant articles will be included. The search strings were the following:

*(ingroup OR intergroup) AND empathy AND (bias OR gap)*

*parochial AND empathy*

*empathy AND bias AND (race OR gender OR sex OR ethnicity)*

*empathy AND bias AND (nationality OR 'socioeconomic status' OR sexuality OR 'sport fans')*

As we aimed to analyse empirical studies of empathy, we pre-selected only journal articles, conference abstracts, and short communications (i.e. we automatically excluded non-empirical records).

Google Scholar on the other hand offers only the choice to limit your search to the article title, which would certainly exclude some relevant records in our opinion. On the other hand, full texts search returns an enormous number of articles (e.g. ~ 45000 records for *intergroup AND empathy AND bias* only!) which is both unrealistic and unneccessary to screen. Therefore, we allowed the search terms to appear anywhere in the text, but they had to appear next to one another, i.e. as a phrase. The search strings were the following:

*"ingroup|intergroup empathy bias|gap"*

*"parochial empathy"*

*"empathy bias" race|gender|sex|ethnicity|nationality|socioeconomic|sexuality|"sport fans"*

As Google Scholar does not enable us to download the list of articles, we used an external software (PublishOrPerish; Harzing, 2007) for scraping the search results.

PsycInfo also allows the users to search for specific strings in the title, abstract, or keywords separately. However, as PsycInfo is specialised in behavioural sciences and as the preliminary searches returned a reasonable number of results, we decided to allow the search terms to appear anywhere in the text. The search strings were as follows:

*(Any Field: ingroup OR Any Field: intergroup) AND Any Field: empathy AND (Any Field: bias OR Any Field: gap)*

*(Any Field: parochial AND Any Field: empathy)*

*(Any Field: race OR Any Field: gender OR Any Field: sex OR Any Field: ethnicity OR Any Field: nationality OR Any Field: "socioeconomic status" OR Any Field: sexuality OR Any Field: "sport fan*")*

Additionally, we pre-selected journal articles (including first postings) and dissertations, thereby excluding books, chapters, comments/replies, etc.

Appendix B

Prisma2020 diagrams per database

Figure B1

*Prisma diagram for Science Direct*

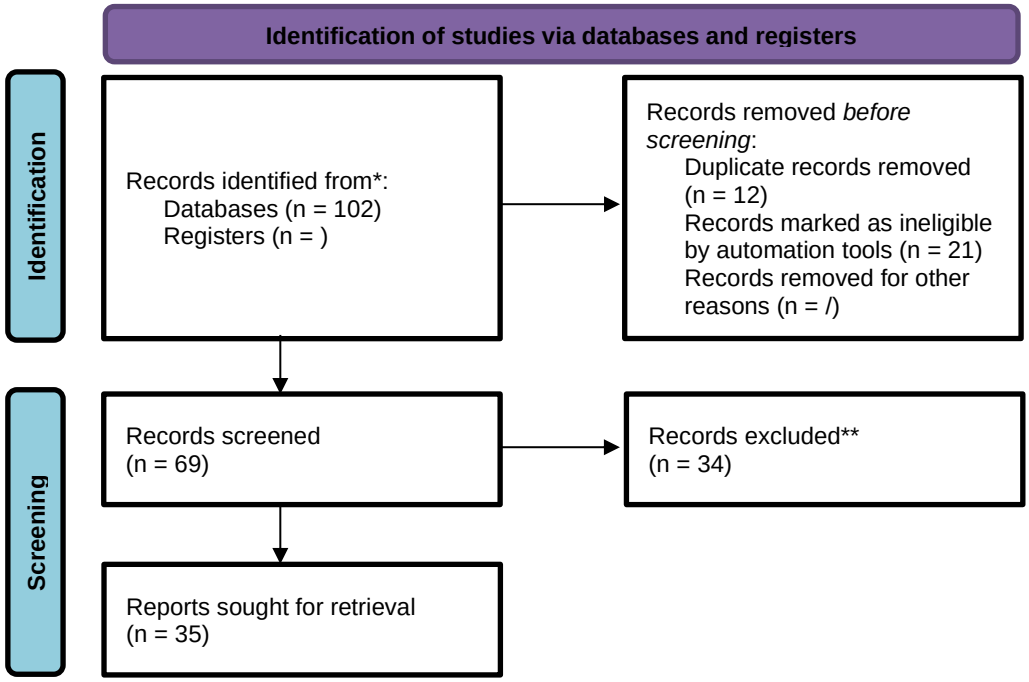


Figure B2

*Prisma diagram for PsycNet*

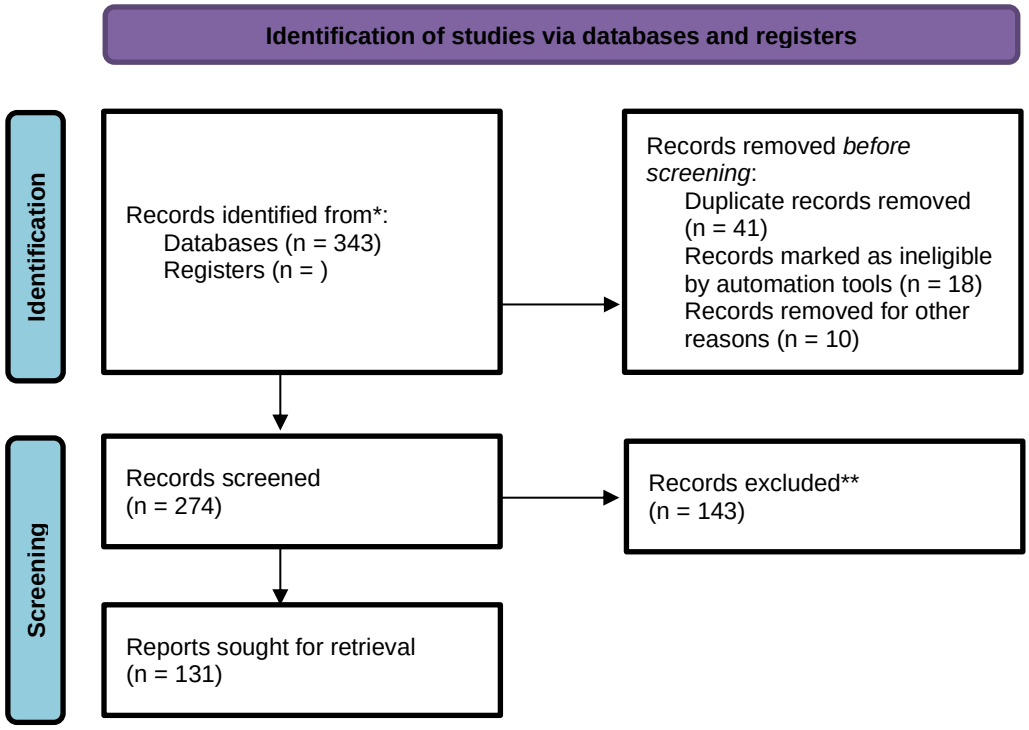
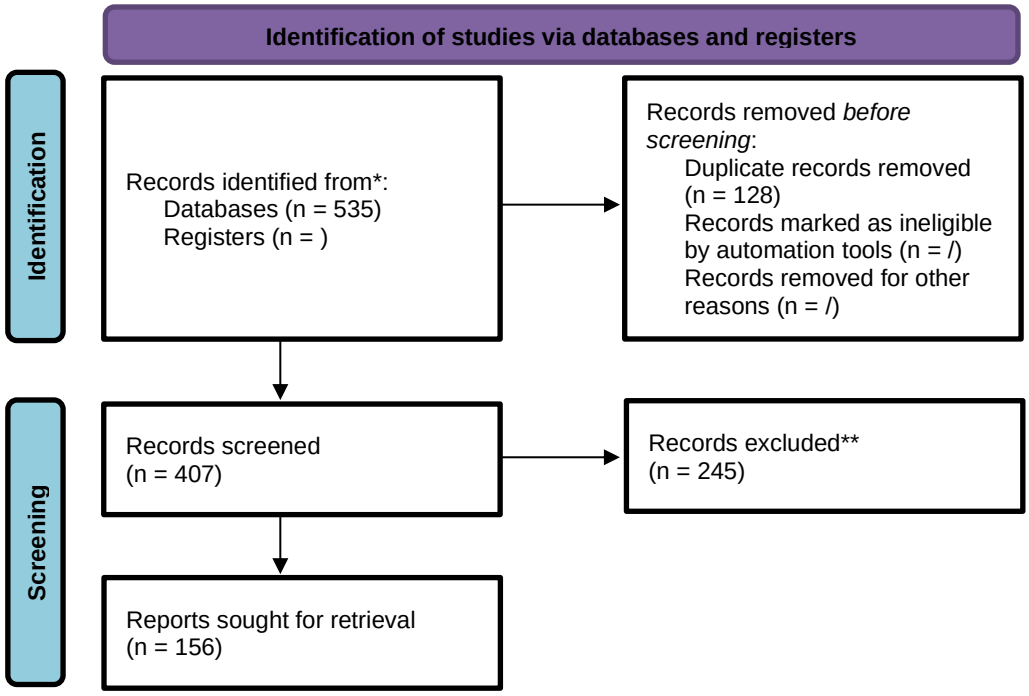



Figure B3

*Prisma diagram for Google Scholar*



Appendix C

## Experiments 1-3: Stimuli

Stimuli used in Experiment 1, Experiment 2, and Experiment 3 were created for this thesis. Photographs for Experiment 1 and Experiment 2 and the video recordings for Experiment 3 were created in one filming session. For Experiment 3, the painful events retold in the videos were pretested. A full set of stimuli as well as pre-test data are uploaded on OSF (https://osf.io/gd5c8/).

## Painful events: Pre-test

Twelve first-person sentences describing painful events were created for Experiment 3. The painful events described were non-life threatening, everyday events that could have happened to anyone, and were partially inspired by the personal experience of the author as well as internet fail videos. Each sentence had a fan-identity salient and neutral version, differing only in the words/phrases signaling the context of the painful event. The descriptions contained 20-29 words. All sentences were phrased informally as if the subject was retelling the event to a friend.

Thirty-four volunteers aged 24-57 ($M \approx 35$) were presented with 6 fan-identity salient and 6 neutral versions of events and asked to rate how *sad, disgusting, surprising, threatening, cheerful, angering, scary, painful,* and *believable* the events were, on a 7-point Likert scale (*not at all – very much*). They were also allowed (and asked) to comment on the content of events.

Summary descriptive statistics for all ratings are presented in Table C1. The highest average rating was for *painful*, closely followed by *surprising*, which is not surprising for the type of events described. Four out of twelve events were rated as slightly more surprising than painful. Two of these events were below 5 on painfulness and they were modified to stress the painful as compared to surprising nature of the event. In addition, one event was modified based on an insightful comment from a participant regarding the anticipated consequences of the event (i.e. that the accident would almost certainly result in a fracture). Painfulness ratings were correlated only to anger ratings ($r = .678$, $p < .05$), with more painful events assessed as more angering. The final set of sentences is presented in Table C2.

Table C1

*Descriptive parameters for sentences describing painful events*

| | M | Max | Min | Diff |
|---|---|---|---|---|
| sad | 3.11 | 3.68 | 2.29 | 1.39 |
| disgusting | 3.24 | 4.26 | 2.39 | 1.87 |
| surprising | 5.42 | 5.77 | 5.19 | 0.58 |
| threatening | 3.83 | 4.61 | 3 | 1.61 |
| cheerful | 1.96 | 2.94 | 1.35 | 1.59 |
| angering | 4.91 | 5.39 | 4.23 | 1.16 |
| scary | 3.8 | 4.39 | 3.29 | 1.1 |
| painful | 5.47 | 5.9 | 4.81 | 1.09 |
| believable | 4.68 | 5.16 | 4.19 | 0.97 |

*Note*. M = Mean; Max = Maximum; Min = Minimum; Diff = Max – Min

Table C2

*Sentences describing painful events*

| Serbian (original) | English (translation) |
|---|---|
| Krenem prošli vikend ***na utakmicu / u prodavnicu***, idem trotoarom i zgazim na nešto što mi probode patiku, podignem nogu, a ono viri ekser. | Last weekend I was going ***to the game /to the store*** and as I was walking the pavement, I stepped onto something that pierced my sneakers. I lifted my foot and I saw a protruding nail. |
| Pre neko veče gledamo ***utakmicu / filmove*** kod Marka i ja da se podignem da ne spadnem sa fotelje i lupim glavom o otvoren prozor. | A few nights ago we were ***watching a game / a film*** at Marko's and as I lifted myself up to avoid sliding from the sofa I banged my head on the opened window |
| Čekam u redu za ***tekmu / banku*** i ne vidim metalni stub ispred sebe i kako sam koraknuo napred šutnem u njega. | I was waiting in line to enter the ***game / bank*** and I didn't see a metal pole in front of me, and as I stepped forward I kicked it. |
| Vraćam se sa ***proslave titule / svadbe*** i pošto je u drvorištu mrkli mrak sapletem se o neku kantu i upadnem u ruže pored staze. | I was returning home from the ***title celebration / wedding***, and as it was pitch black in the yard I tripped over a bucket and fell into the roses next to the path. |
| Bio sam sa drugarima iz ***kluba / srednje*** u kafiću i krenem do toaleta, okliznem se na mokar pod i lupim glavom o ivicu stola.[1] | I was in a bar with my ***club / high school*** and as I was walking to the bathroom, I slipped on the wet floor and hit my head on the table edge. |
| Pre neki dan trčim da stignem do ***stadiona / tržnog centra*** pre nego što zatvore ***blagajnu / apoteku*** okliznem se na mokar pešački prelaz, padnem i razbijem lakat.[1] | Some days ago I was running to make it to the ***stadium / mall*** before the ***ticket office / pharmacy*** closed, and I slipped on a wet zebra and smashed my elbow. |
| Na prošlom ***derbiju / koncertu*** na kom sam bio na samom početku krene guranje i čovek iza mene mi slučajno prospe vrelu kafu po leđima. | On the last ***derby / concert*** I visited, not long after the start there was some pushing and the man behind me accidentally spilled hot coffee on my back. |
| Vraćam se biciklom sa ***skupštine kluba / nekog sastanka*** i cimnem volanom da izbegnem rupu na putu, zakačim granu od drveta i padnem koliko sam dug. | I was riding my bike from ***a fan club meeting /some meeting*** and as I swung the wheel to avoid a hole in the road, I hit a tree branch and fell flat on my face. |
| Na ***poluvremenu / pauzi na nekom seminaru*** krenem da kupim vodu i kako sam ustajao zakačim o polomljenu stolicu ***na tribinama*** / ***Ø*** i isečem se po butini. | During ***the half-time break / a break on a seminar*** I wanted to buy water and I was standing up I grazed a broken chair and cut my thigh. |
| Krenuo da kupim kartu za ***utakmicu / svirku*** i okrenem glavu jer me neko pozvao a nastavljam da hodam i zakucam se u uličnu rasvetu. | I was walking to buy the tickets ***for the game /for the gig***, and I turned my head around because someone called me while continuing to walk and I smashed into the street light. |
| Treba da krenem kod drugara da gledamo ***utakmicu kupa / neki dokumentarac***, gladan a žurim, i dok sam u brzini sekao hleb isečem se po palcu. | I was getting ready to go to a friend's to watch ***the game / a documentary***, hungry and late, and I was hurrying to slice the bread I cut my thumb. |
| U povratku sa ***utakmice / rođendana*** ugasi mi se auto i ja pod haubom da vidim šta je, otkači se držač haube i ona padne i udari me po sred potiljka. | While getting back from the ***game / birthday party*** my car stopped, and as I opened the hood to see what was it, the hood holder failed and it hit me right in the back of my head. |

*Note*. [1] – Events modified after the pretest. Phrases indicating the fan-identity salient or neutral context are italicised.

Filming session

Twelve young male volunteers (aged 20-28, M = 22) were photographed and filmed in individual sessions. After providing informed consent, they chose an event they were going to retell as if it had happened to them (for Experiment 3) and they were given time to learn it and rehearse it. After they indicated they were ready, models were seated on a bench in front of a black background approximately 1 m from the camera (Canon EOS 600d) on a tripod. They were given a Red Star, Partizan, or Vojvodina jersey to dress. Then they were photographed chest-up:

1) While posing with a neutral facial expression (Experiment 1)

- The models were instructed to imagine they were photographed for their identity card

2) While posing with a painful facial expression (Experiment 2)

- The models were instructed to imagine they had stepped on a Lego or stubbed their pinky on a chair and to make a painful facial expression signaling pain intensity 8 on a scale of 1-10 (the instructions were adapted from the Delaware pain database, Mende-Siedlecki, Qu-Lee, Lin, Drain, & Goharzad, 2020)

3) Filmed while retelling two versions of the painful event they had previously chosen. (Experiment 3).

This sequence was repeated 3 times, i.e. once with each jersey. The order of the jerseys and the salience of the events was counterbalanced (Latin square) to account for practice effects.

Photographs for Experiment 1 and Experiment 2 were chosen by two independent raters. The chosen photographs (two for each model – one with neutral and one with painful expression) were digitally edited with Adobe Photoshop. Each photograph was cropped to 3584 x 3456 px. A solid black background was added to each photograph and three different jerseys were digitally added to the same neutral and painful photo of each model (6 photos per model). On neutral expressions, needles, and q-tips were added on the model's left or right cheek for each jersey (3 jerseys x 4 photos = 12 photos per model). Although we photographed the models in three different jerseys, we decided to digitally add them to the photo because we wanted to keep the expressions constant instead of pretesting several photos of each participant in each jersey and finding equal pairs by the intensity and various hypothetical confounding variables. Videos for Experiment 3 were cropped just before and after the stimulus sentence and background noise was removed. For video editing, we used OpenShot video editor and Audacity.

Appendix D

Sport Spectator Identification Scale (SSIS) reliability analysis

We analysed the SSIS on the full sample of participants that participated in all experiments (N = 532).

Table D1

*Descriptive parameters for SSIS items*

| item | Min | Max | M | SD | zSk | zKu |
|---|---|---|---|---|---|---|
| U kojoj meri ti je važno da navedeni tim pobedi? | 1 | 8 | 6.39 | 1.82 | -10.67 | 2.64 |
| U kojoj meri smatraš sebe navijačem navedenog tima? | 1 | 8 | 5.93 | 1.91 | -6.77 | -2.04 |
| U kojoj meri te prijatelji smatraju navijačem navedenog tima? | 1 | 8 | 5.89 | 2.03 | -7.06 | -2.15 |
| Tokom sezone, koliko pomno pratiš navedeni tim putem bilo kojeg od sledećih medija: uživo ili na televiziji, preko radija, ili preko televizijskih ili novinskih izveštaja | 1 | 8 | 5.73 | 2.09 | -5.26 | -4.11 |
| U kojoj meri ti je važno to što si navijač navedenog tima? | 1 | 8 | 5.49 | 2.44 | -5.20 | -5.00 |
| U kojoj meri mrziš najvećeg rivala navedenog tima? | 1 | 8 | 3.86 | 2.62 | 3.61 | -6.39 |
| Koliko često ističeš ime ili obeležja navedenog tima na poslu, kod kuće ili na svojoj odeći? | 1 | 8 | 3.53 | 2.23 | 4.92 | -4.40 |

*Note*. Min – Minimum; Max – Maximum; M – Mean; SD – Standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis

Table D2

*Scale reliability: 7 items (full scale)*

| Cronbach's Alpha | Cronbach's Alpha Based on Standardised Items | N of Items |
|---|---|---|
| 0.903 | 0.911 | 7 |

Table D3

*Item-total statistics: 7 items*

| item | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|
| U kojoj meri ti je važno da navedeni tim pobedi? | 0.770 | 0.885 |
| U kojoj meri smatraš sebe navijačem navedenog tima? | 0.831 | 0.878 |
| U kojoj meri te prijatelji smatraju navijačem navedenog tima? | 0.811 | 0.879 |
| Tokom sezone, koliko pomno pratiš navedeni tim putem bilo kojeg od sledećih medija: uživo ili na televiziji, preko radija, ili preko televizijskih ili novinskih izveštaja | 0.761 | 0.884 |
| U kojoj meri ti je važno to što si navijač navedenog tima? | 0.800 | 0.878 |
| U kojoj meri mrziš najvećeg rivala navedenog tima? | 0.516 | 0.917 |
| Koliko često ističeš ime ili obeležja navedenog tima na poslu, kod kuće ili na svojoj odeći? | 0.611 | 0.900 |

Table D4

*Scale reliability: 6 items*

| Cronbach's Alpha | Cronbach's Alpha Based on Standardised Items | N of Items |
|---|---|---|
| .917 | .921 | 6 |

Table D5

*Item-total statistics: 6 items*

| item | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|
| U kojoj meri ti je važno da navedeni tim pobedi? | 0.780 | 0.901 |
| U kojoj meri smatraš sebe navijačem navedenog tima? | 0.853 | 0.891 |
| U kojoj meri te prijatelji smatraju navijačem navedenog tima? | 0.832 | 0.892 |
| Tokom sezone, koliko pomno pratiš navedeni tim putem bilo kojeg od sledećih medija: uživo ili na televiziji, preko radija, ili preko televizijskih ili novinskih izveštaja | 0.784 | 0.899 |
| U kojoj meri ti je važno to što si navijač navedenog tima? | 0.805 | 0.897 |
| Koliko često ističeš ime ili obeležja navedenog tima na poslu, kod kuće ili na svojoj odeći? | 0.583 | 0.928 |

Table D6

*Scale reliability: 5 items*

| Cronbach's Alpha | Cronbach's Alpha Based on Standardised Items | N of Items |
|---|---|---|
| 0.928 | 0.931 | 5 |

Table D7

*Item-total statistics: 5 items*

| item | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|
| U kojoj meri ti je važno da navedeni tim pobedi? | 0.793 | 0.916 |
| U kojoj meri smatraš sebe navijačem navedenog tima? | 0.866 | 0.902 |
| U kojoj meri te prijatelji smatraju navijačem navedenog tima? | 0.841 | 0.906 |
| Tokom sezone, koliko pomno pratiš navedeni tim putem bilo kojeg od sledećih medija: uživo ili na televiziji, preko radija, ili preko televizijskih ili novinskih izveštaja | 0.788 | 0.916 |
| U kojoj meri ti je važno to što si navijač navedenog tima? | 0.800 | 0.918 |

Appendix E

Sample description

Table E1

*Experiments 1-3: Age and socioeconomic status*

| | M | Mdn | SD | Min | Max | Q75 |
|---|---|---|---|---|---|---|
| age | 24.55 | 21.00 | 8.92 | 17 | 63 | 24.75 |
| SES (1-101) | 58.68 | 57.50 | 17.67 | 14 | 101 | 70.00 |

*Note*. SES – Socioeconomic status; M – Mean; Mdn – Median; SD – Standard deviation; Min – Minimum; Max – Maximum; Q75 – 75$^{th}$ percentile (Third Quartile); N = 204

Table E2

*Experiments 1-3: Education and employment*

| Highest degree obtained | |
|---|---|
| | Frequency (%) |
| primary | 3 (1.5) |
| secondary | 148 (72.5) |
| post-secondary | 19 (9.3) |
| BA/BSc | 19 (9.3) |
| MA/MSc | 11 (5.4) |
| Magisterium | 2 (1.0) |
| PhD | 2 (1.0) |
| **Currently a student** | |
| yes | 142 (69.7) |
| no | 62 (30.4) |

| Employment status | | |
|---|---|---|
| | | Frequency (%) |
| unemployed | not checked | 107 (52.5) |
| | checked | 97 (47.5) |
| occasionally employed | not checked | 165 (80.9) |
| | checked | 39 (19.1) |
| temporarily employed | not checked | 183 (89.7) |
| | checked | 21 (10.3) |
| full-time employed | not checked | 157 (77.0) |
| | checked | 47 (23.0) |
| self-employed | not checked | 191 (93.6) |
| | checked | 13 (6.4) |

*Note*. For employment status, participants were allowed to check more than one option.

Table E3

*Experiments 1-3: Fan identity-related assessments by Red Star and Partizan fans*

| | Partizan fans | | Red Star fans | |
|---|---|---|---|---|
| | Answer, Frequency (%) | | Answer, Frequency (%) | |
| greatest rival | Red Star, 76 (98.7) | Other, 1 (1.3) | Partizan, 119 (93.7) | Other, 8 (6.3) |
| second best rival | Vojvodina, 49 (63.6) | Other, 28 (36.4) | Vojvodina, 59 (46.5) | Other, 68 (53.5) |
| fan group membership | yes, 6 (7.8) | no, 71 (92.2) | yes, 16 (12.6) | no, 111 (87.4) |

*Note*. Partizan fans N = 77; Red Star fans N = 127.

Table E4

*Experiments 1-3: Fan identity-related assessments by Red Star and Partizan fans*

| Question | | Likert scale poles (1-5) | Partizan fans M | Partizan fans SD | Red Star fans M | Red Star fans SD |
|---|---|---|---|---|---|---|
| How would you you're your team's relative status compared to the greatest rival… | …presently? | we are the worse team – we are the better team | 2.79 | 1.30 | 4.61 | .71 |
| | ….historically? | | 3.53 | 1.19 | 4.79 | .60 |
| How would you you're your team's relative status compared to the second best rival… | …presently? | | 4.60 | .89 | 4.83 | .55 |
| | …historically? | | 4.77 | .71 | 4.94 | .38 |
| How frequently do you… | …watch live broadcasts or highlights of football matches? | never – almost every day/always | 3.95 | 1.06 | 4.02 | .94 |
| | … follow your team games on home ground? | | 2.77 | 1.36 | 2.93 | 1.27 |
| | … follow your team when they play away in the country? | | 2.08 | 1.17 | 2.13 | 1.33 |
| | … follow your team when they play away in foreign countries? | | 1.65 | 1.13 | 1.92 | 1.35 |
| How similar do you think you are to… | …own team fans? | not at all similar – very similar | 2.95 | 1.13 | 2.97 | 1.11 |
| | …rival team fans | | 1.79 | .86 | 2.06 | 1.09 |
| How different do you think you are from… | …own team fans? | not at all different – very different | 3.10 | 1.11 | 3.10 | 1.09 |
| | …rival team fans | | 3.91 | 1.15 | 3.78 | 1.18 |

*Note*. Partizan fans N = 77; Red Star fans N = 127.

Table E5

*Experiment 4: Age and socioeconomic status*

| | M | Mdn | SD | Min | Max | Q75 |
|---|---|---|---|---|---|---|
| age | 27.42 | 23.00 | 10.69 | 16 | 68 | 35.00 |
| SES (1-101) | 57.76 | 56.00 | 17.55 | 1 | 101 | 70.25 |

*Note*. SES – Socioeconomic status; ; M – Mean; Mdn – Median; SD – Standard deviation; Min – Minimum; Max – Maximum; Q75 – 75$^{th}$ percentile (Third Quartile); N = 140

Table E6

*Experiment 4: Education and employment*

| Highest degree obtained | |
|---|---|
| | Frequency (%) |
| primary | 1 (0.6) |
| secondary | 107 (67.7) |
| post-secondary | 8 (5.1) |
| BA/BSc | 32 (20.3) |
| MA/MSc | 7 (4.4) |
| Magisterium | 1 (0.6) |
| PhD | 2 (1.3) |
| Currently a student | |
| yes | 97 (61.4) |
| no | 61 (38.6) |

| Employment status | | |
|---|---|---|
| | | Frequency (%) |
| unemployed | not checked | 82 (51.9) |
| | checked | 76 (48.1) |
| occasionally employed | not checked | 135 (85.4) |
| | checked | 23 (14.6) |
| temporarily employed | not checked | 135 (85.4) |
| | checked | 23 (14.6) |
| full-time employed | not checked | 127 (80.4) |
| | checked | 31 (19.6) |
| self-employed | not checked | 145 (91.8) |
| | checked | 13 (8.2) |

*Note*. For employment status, participants were allowed to check more than one option.

Table E7

*Experiments 1-3: Fan identity-related assessments by Red Star and Partizan fans*

| | Partizan fans | | Red Star fans | |
|---|---|---|---|---|
| | Answer, Frequency (%) | | Answer, Frequency (%) | |
| greatest rival | Red Star, 75 (100) | Other, 0 (0) | Partizan, 79 (95.2) | Other, 4 (4.8) |
| second best rival | Vojvodina, 43 (57.3) | Other, 32 (42.7) | Vojvodina, 42 (50.6) | Other, 41 (49.4) |
| fan group membership | yes, 6 (8.0) | no, 69 (92.0) | yes, 2 (2.4) | no, 81 (97.6) |

Note. Partizan fans N = 75; Red Star fans N = 83.

Table E8

*Experiments 1-3: Fan identity-related assessments by Red Star and Partizan fans*

| Question | | Likert scale poles (1-5) | Partizan fans | | Red Star fans | |
|---|---|---|---|---|---|---|
| | | | M | SD | M | SD |
| How would you you're your team's relative status compared to the greatest rival… | …presently? | we are the worse team – we are the better team | 2.08 | 1.06 | 4.81 | .45 |
| | ….historically? | | 3.23 | 1.17 | 4.77 | .48 |
| How would you you're your team's relative status compared to the second best rival… | …presently? | | 4.16 | 1.15 | 4.90 | .43 |
| | …historically? | | 4.68 | .72 | 4.96 | .24 |
| How frequently do you… | …watch live broadcasts or highlights of football matches? | never – almost every day/always | 4.07 | .92 | 4.12 | 1.05 |
| | … follow your team games on home ground? | | 2.51 | 1.20 | 2.65 | .97 |
| | … follow your team when they play away in the country? | | 1.71 | .96 | 1.77 | 1.00 |
| | … follow your team when they play away in foreign countries? | | 1.48 | .95 | 1.57 | 1.07 |
| How similar do you think you are to… | …own team fans? | not at all similar – very similar | 2.80 | .97 | 2.66 | .99 |
| | …rival team fans | | 1.80 | .92 | 1.96 | .97 |
| How different do you think you are from… | …own team fans? | not at all different – very different | 3.27 | .89 | 3.27 | .96 |
| | …rival team fans | | 4.15 | .90 | 3.90 | 1.10 |

Note. Partizan fans N = 75; Red Star fans N = 83.

Appendix F

Additional descriptive statistics for Experiment 1 and Experiment 2

Table F1

*Experiment 1: Descriptive statistics for pain intensity and self-unpleasantness ratings*

| rating | | | Min | Max | M | SD | zSk | zKu | P90 |
|---|---|---|---|---|---|---|---|---|---|
| pain intensity | painful | ingroup | 1.00 | 7.00 | 2.66 | 1.50 | 3.902 | -0.298 | 5.000 |
| | | rival outgroup | 1.00 | 7.00 | 2.72 | 1.52 | 3.989 | -0.215 | 5.175 |
| | | neutral outgroup | 1.00 | 6.13 | 2.65 | 1.48 | 3.963 | -0.325 | 5.175 |
| | neutral | ingroup | 1.00 | 6.25 | 1.60 | 0.92 | 12.074 | 7.448 | 2.900 |
| | | rival outgroup | 1.00 | 7.00 | 1.65 | 0.97 | 12.294 | 8.174 | 3.000 |
| | | neutral outgroup | 1.00 | 5.50 | 1.57 | 0.86 | 10.826 | 5.376 | 2.900 |
| self-unpleasantness | painful | ingroup | 1.00 | 7.00 | 2.53 | 1.65 | 5.184 | 0.108 | 5.000 |
| | | rival outgroup | 1.00 | 7.00 | 2.72 | 1.83 | 4.772 | -0.252 | 6.000 |
| | | neutral outgroup | 1.00 | 7.00 | 2.62 | 1.75 | 5.181 | 0.011 | 5.875 |
| | neutral | ingroup | 1.00 | 6.25 | 1.53 | 0.95 | 12.540 | 7.072 | 2.875 |
| | | rival outgroup | 1.00 | 7.00 | 1.74 | 1.31 | 12.416 | 6.486 | 3.375 |
| | | neutral outgroup | 1.00 | 6.25 | 1.56 | 1.04 | 13.093 | 7.549 | 3.000 |

*Note*. Min – Minimum; Max – Maximum; M – Mean; SD – Standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; P90 – 90$^{th}$ percentile; N = 147

Table F2

*Experiment 2: Descriptive statistics for pain intensity and self-unpleasantness ratings*

| rating | | | Min | Max | M | SD | zSk | zKu | P90 |
|---|---|---|---|---|---|---|---|---|---|
| pain intensity | painful | ingroup | 1.00 | 6.75 | 4.11 | 1.24 | -1.596 | -0.258 | 5.750 |
| | | rival outgroup | 1.00 | 7.00 | 4.25 | 1.36 | -1.720 | -0.447 | 6.000 |
| | | neutral outgroup | 1.00 | 6.75 | 4.18 | 1.26 | -0.491 | -0.722 | 5.750 |
| | neutral | ingroup | 1.00 | 4.50 | 1.42 | 0.64 | 9.621 | 4.232 | 2.250 |
| | | rival outgroup | 1.00 | 7.00 | 1.61 | 1.01 | 12.611 | 8.227 | 3.000 |
| | | neutral outgroup | 1.00 | 5.75 | 1.56 | 0.88 | 11.749 | 6.850 | 2.500 |
| self-unpleasantness | painful | ingroup | 1.00 | 6.50 | 2.61 | 1.58 | 3.309 | -0.711 | 5.225 |
| | | rival outgroup | 1.00 | 7.00 | 2.84 | 1.85 | 3.421 | -0.746 | 5.975 |
| | | neutral outgroup | 1.00 | 6.75 | 2.76 | 1.76 | 3.387 | -0.798 | 5.725 |
| | neutral | ingroup | 1.00 | 4.25 | 1.33 | 0.61 | 11.228 | 5.221 | 2.250 |
| | | rival outgroup | 1.00 | 7.00 | 1.71 | 1.39 | 11.909 | 5.433 | 3.925 |
| | | neutral outgroup | 1.00 | 6.50 | 1.50 | 0.93 | 12.347 | 7.341 | 2.750 |

Note. Min – Minimum; Max – Maximum; M – Mean; SD – Standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; P90 – 90$^{th}$ percentile; N = 140

Average ratings in both experiments were winsorized by using median absolute deviation (MAD), a robust measure of variability, insensitive to both sample size and outliers (Leys, Ley, Klein, Bernard, & Licata, 2012). We replaced all the values 2.5x MAD above or below the median with that value. As presented in Table G1, the winsorization did not change painful stimuli average ratings as compared to raw data, but 7/12 neutral stimuli ratings were fixed to a constant.

Table F3

*Descriptive statistics for winsorized ratings in Experiment 1 and Experiment 2*

| | rating | Min | Max | M | SD | zSk | zKu | P90 |
|---|---|---|---|---|---|---|---|---|
| Experiment 1 – pain intensity | | | | | | | | |
| painful | ingroup | 1.00 | 6.42 | 2.66 | 1.49 | 3.729 | -1.088 | 5.000 |
| | rival outgroup | 1.00 | 6.54 | 2.72 | 1.51 | 3.861 | -0.783 | 5.175 |
| | neutral outgroup | 1.00 | 5.96 | 2.64 | 1.47 | 3.888 | -0.927 | 5.175 |
| neutral | ingroup | 1.00 | 1.59 | 1.27 | 0.27 | 1.061 | -4.610 | 1.588 |
| | rival outgroup | 1.00 | 2.18 | 1.45 | 0.48 | 2.576 | -3.605 | 2.177 |
| | neutral outgroup | 1.00 | 2.18 | 1.41 | 0.46 | 3.539 | -2.883 | 2.177 |
| Experiment 1 – self-unpleasantness | | | | | | | | |
| painful | ingroup | 1.00 | 5.71 | 2.48 | 1.54 | 1.612 | 0.956 | 5.000 |
| | rival outgroup | 1.00 | 5.71 | 2.63 | 1.66 | 1.588 | 1.042 | 5.707 |
| | neutral outgroup | 1.00 | 5.71 | 2.55 | 1.60 | 1.592 | 1.005 | 5.707 |
| neutral | ingroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| | rival outgroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| | neutral outgroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| Experiment 2 – pain intensity | | | | | | | | |
| painful | ingroup | 1.47 | 6.75 | 4.12 | 1.22 | -1.122 | -1.196 | 5.750 |
| | rival outgroup | 1.00 | 7.00 | 4.25 | 1.36 | -1.720 | -1.098 | 6.000 |
| | neutral outgroup | 1.00 | 6.75 | 4.18 | 1.26 | -0.491 | -1.774 | 5.750 |
| neutral | ingroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| | rival outgroup | 1.00 | 2.18 | 1.39 | 0.47 | 3.637 | -2.755 | 2.177 |
| | neutral outgroup | 1.00 | 2.18 | 1.40 | 0.47 | 3.385 | -2.840 | 2.177 |
| Experiment 2 – self-unpleasantness | | | | | | | | |
| painful | ingroup | 1.00 | 6.50 | 2.61 | 1.58 | 3.309 | -1.746 | 5.225 |
| | rival outgroup | 1.00 | 6.88 | 2.84 | 1.85 | 3.373 | -1.909 | 5.975 |
| | neutral outgroup | 1.00 | 5.71 | 2.71 | 1.67 | 2.778 | -2.733 | 5.686 |
| neutral | ingroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| | rival outgroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |
| | neutral outgroup | 1.00 | 1.00 | 1.00 | / | / | / | 1.000 |

Note. Min – Minimum; Max – Maximum; M – Mean; SD – Standard deviation; zSk – standardised Skewness; zKu – standardised Kurtosis; P90 – 90$^{th}$ percentile

Appendix G

Supplementary analyses for Experiment 1 and Experiment 2

Pain-neutral difference scores for raw data

We calculated pain-neutral difference scores on raw data to account for non-normal distributions. We analysed ingroup, rival outgroup and neutral outgroup rating differences with separate ANOVAs for pain intensity and self-unpleasantness separately, in both experiments.

Experiment 1

For pain intensity pain-neutral difference in ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.790 $\chi^2(2) = 34.27$, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .826. The main effect of group was not significant, $F(1.65, 241.24) = .05$, $p$ = .929, indicating that the pain-neutral differences in pain intensity ratings were similar across group identity of the target.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.746, $\chi^2(2) = 42.42$, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .798. The main effect of group was not significant, $F(1.60, 232.92) = 2.17$, $p$ = .128, indicating that the pain-neutral differences in self-unpleasantness ratings were similar across group identity of the target.

Experiment 2

For pain intensity pain-neutral difference in ratings, Mauchly's W was not significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.991 $\chi^2(2) = 1.24$, $p$ = .539), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .991. The main effect of group was not significant, $F(1.98, 275.54) = .45$, $p$ = .448, indicating that the pain-neutral differences in pain intensity ratings were similar across group identity of the target.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.946, $\chi^2(2) = 7.61$, $p$ = .022), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .949. The main effect of group was not significant, $F(1.90, 263.84) = 1.78$, $p$ = .173, indicating that the pain-neutral differences in self-unpleasantness ratings were similar across group identity of the target.

Painful stimuli: One-way ANOVA

We compared pain intensity and self-unpleasantness ratings on painful stimuli only, for both Experiments.

Experiment 1

For pain intensity ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.835, $\chi^2(2)$ = 26.20, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .858. The main effect of group was not significant, $F(1.72, 250.58)$ = 1.28, $p$ = .276, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.809, $\chi^2(2)$ = 30.65, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .840. The main effect of group was significant, $F(1.68, 245.27)$ = 3.41, $p$ = .043, indicating that participants rated the self-unpleasantness elicited by painful images of ingroups, rival outgroups, and neutral outgroups differently. However, no post-hoc Bonferroni-corrected comparison was significant (ingroup – rival outgroup Mean difference = -.186, $p$ = .092; ingroup – neutral outgroup mean difference = -.088, $p$ = .538; rival outgroup – neutral outgroup Mean difference = .098, $p$ = .328).

Experiment 2

For pain intensity ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.867, $\chi^2(2)$ = 19.71, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .883. The main effect of group was not significant, $F(1.77, 245.31)$ = 1.50, $p$ = .227, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.829, $\chi^2(2)$ = 25.85, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .854. The main effect of group was significant, $F(1.71, 237.44)$ = 3.53, $p$ = .038, indicating that participants rated the self-unpleasantness elicited by painful images of ingroups, rival outgroups, and neutral outgroups differently. However, no post-hoc Bonferroni-corrected comparison was significant (ingroup – rival outgroup Mean difference = -.225, $p$ = .086; ingroup – neutral outgroup mean difference = -.143, $p$ = .224; rival outgroup – neutral outgroup Mean difference = .082, $p$ = .796).

Painful stimuli: Friedman test

We compared ingroup, rival outgroup and neutral outgroup pain intensity and self-unpleasantness ratings for painful stimuli only with non-parametric Friedman test. The results are presented in table below. No comparison was significant.

Table G1

*Friedman test parameters for comparison of ingroup, rival outgroup and neutral outrgroup ratings for painful stimuli*

| | | $\chi^2$(df) | $p$ |
|---|---|---|---|
| Experiment 1 | pain intensity | $\chi^2(2)$ = .990 | .610 |
| | self-unpleasantness | $\chi^2(2)$ = .487 | .784 |
| Experiment 2 | pain intensity | $\chi^2(2)$ = 1.71 | .426 |
| | self-unpleasantness | $\chi^2(2)$ = 4.09 | .129 |

Painful stimuli – winsorized data

We compared ingroup, rival outgroup and neutral outgroup pain intensity and self-unpleasantness ratings for painful stimuli only, on previously winsorized data.

Experiment 1

For pain intensity ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.859, $\chi^2(2)$ = 22.04, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .876. The main effect of group was not significant, $F(1.75, 255.92) = 1.40$, $p$ = .248, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.793, $\chi^2(2)$ = 33.70, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .828. The main effect of group was not significant, $F(1.66, 241.85) = 3.00$, $p$ = .061, indicating that participants rated the self-unpleasantness elicited by painful images of ingroups, rival outgroups, and neutral outgroups similarly on average.

Experiment 2

For pain intensity ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.880, $\chi^2(2)$ = 17.65, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .893. The main effect of group was not significant, $F(1.78, 248.20) = 1.31$, $p$ = .269, indicating that participants rated the pain intensity of ingroups, rival outgroups, and neutral outgroups similarly on average.

For self-unpleasantness ratings, Mauchly's W was significant for the main effect of group (ingroup, rival outgroup, or neutral outgroup) ($W$ =.797, $\chi^2(2)$ = 31.24, $p$ = .000), and the Greenhouse-Geisser estimate of the deviation from sphericity was $\varepsilon$ = .832. The main effect of group was significant, $F(1.66, 231.74) = 3.55$, $p$ = .038, indicating that participants rated the self-unpleasantness elicited by painful images of ingroups, rival outgroups, and neutral outgroups differently. However, post hoc comparisons indicated that no contrast was significant after Bonferroni correction (ingroup – rival outgroup Mean difference = -222, $p$ = .087; ingroup – neutral outgroup mean difference = -.099, $p$ = .552; rival outgroup – neutral outgroup Mean difference = .123, $p$ = .281).

Appendix H

Full correlation table for Identity measures, trait empathy, and prejudice for both datasets.

Table H1

*Experiment 1-3 dataset*

| | Fan identity | | | Interpersonal Reactivity Index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIS | BIRG-ing | CORF-ing | F | PT | EC | PD | SDO | IAT (\|D\|) |
| SSIS | 1 | .505** | -0.027 | -0.022 | -0.123 | 0.032 | .225** | .157* | -0.037 |
| BIRGing | .496** | 1 | 0.045 | 0.034 | -.227** | -0.136 | .263** | .281** | 0.053 |
| CORFing | -0.031 | 0.069 | 1 | -0.084 | -0.043 | -0.096 | .199** | -0.025 | 0.005 |
| F | -0.016 | 0.032 | -0.110 | 1 | .370** | .363** | -0.059 | -0.083 | -0.018 |
| PT | -0.107 | -.209** | -0.047 | .408** | 1 | .407** | -0.126 | -.308** | -0.126 |
| EC | 0.039 | -.151* | -0.084 | .366** | .437** | 1 | -0.007 | -.383** | -0.018 |
| PD | .226** | .268** | .184** | -0.060 | -0.098 | -0.003 | 1 | .192** | -0.022 |
| SDO | .147* | .295** | -0.045 | -0.098 | -.315** | -.388** | .186** | 1 | 0.133 |
| (\|D\|) | -0.005 | 0.076 | -0.002 | -0.007 | -0.128 | -0.024 | -0.028 | 0.111 | 1 |

Note. Pearson *r* – below the diagonal; Spearman Rho – above the diagonal; SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; |D| – Implicit associations test D measure, absolute value; * $p < .05$; ** $p < .01$; N = 204

Table H2

*Experiment 4 dataset*

| | Fan identity | | | Interpersonal Reactivity Index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIS | BIRG-ing | CORF-ing | F | PT | EC | PD | SDO | \|D\| |
| SSIS | 1 | .508** | -0.033 | .201* | 0.002 | 0.126 | -0.111 | 0.150 | .321** |
| BIRGing | .494** | 1 | 0.155 | .185* | 0.047 | 0.027 | -0.089 | .252** | .173* |
| CORFing | -0.055 | 0.153 | 1 | 0.048 | -0.123 | 0.055 | .208** | 0.015 | -0.092 |
| F | .187* | .179* | 0.097 | 1 | .307** | .446** | 0.128 | -.285** | 0.138 |
| PT | -0.020 | 0.033 | -0.095 | .316** | 1 | .391** | -0.078 | -.209** | 0.001 |
| EC | 0.106 | -0.013 | 0.057 | .456** | .436** | 1 | 0.006 | -.443** | 0.129 |
| PD | -0.129 | -0.081 | .200* | 0.152 | -0.063 | 0.023 | 1 | 0.048 | -0.024 |
| SDO | 0.152 | .258** | -0.025 | -.263** | -.255** | -.465** | 0.026 | 1 | -0.103 |
| (\|D\|) | .310** | 0.163 | -0.063 | .172* | 0.004 | 0.114 | -0.048 | -0.075 | 1 |

Note. Pearson *r* – below the diagonal; Spearman Rho – above the diagonal; SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; |D| – Implicit associations test D measure, absolute value; * $p < .05$; ** $p < .01$; N = 158

Appendix I

Full correlation table for empathy ratings with identity measures, trait empathy, and prejudice for Experiment 3.

Table I1

*Experiment 3: Pain intensity ratings - Full correlation table*

| rating | stimuli | group | Pain intensity | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | salient | | | neutral | | |
| | | | IG | RO | DO | IG | RO | DO |
| pain intensity | salient | IG | 1 | .603** | .586** | .588** | .458** | .511** |
| | | RO | .548** | 1 | .640** | .638** | .551** | .619** |
| | | DO | .592** | .660** | 1 | .502** | .558** | .652** |
| | neutral | IG | .573** | .639** | .523** | 1 | .504** | .555** |
| | | RO | .430** | .588** | .567** | .493** | 1 | .474** |
| | | DO | .477** | .617** | .648** | .566** | .447** | 1 |
| self-unpleasantness | salient | IG | .677** | .448** | .393** | .407** | .393** | .262** |
| | | RO | .450** | .622** | .434** | .432** | .451** | .448** |
| | | DO | .535** | .472** | .628** | .407** | .409** | .449** |
| | neutral | IG | .420** | .483** | .356** | .639** | .384** | .336** |
| | | RO | .420** | .421** | .280** | .412** | .556** | .253* |
| | | DO | .455** | .497** | .501** | .434** | .419** | .637** |
| Schadenfreude | salient | IG | -.233* | -0.151 | -0.103 | -0.120 | -0.109 | -0.088 |
| | | RO | -0.026 | -.235* | -0.119 | -0.008 | -0.183 | -0.138 |
| | | DO | -0.014 | 0.094 | -0.075 | 0.044 | 0.086 | 0.001 |
| | neutral | IG | -0.004 | -0.095 | -0.025 | -0.184 | 0.031 | -0.015 |
| | | RO | -0.033 | -0.160 | -0.156 | 0.019 | -.280** | -0.179 |
| | | DO | -0.023 | -0.007 | -0.014 | 0.100 | -0.034 | -0.027 |
| empathic concern | salient | IG | .658** | .424** | .400** | .375** | .255* | .362** |
| | | RO | .425** | .726** | .565** | .490** | .358** | .552** |
| | | DO | .537** | .422** | .562** | .411** | .245* | .432** |
| | neutral | IG | .406** | .408** | .311** | .570** | 0.197 | .385** |
| | | RO | .394** | .499** | .459** | .445** | .574** | .389** |
| | | DO | .398** | .550** | .459** | .404** | .228* | .627** |
| perspective taking | salient | IG | .417** | .283** | .301** | .206* | .337** | 0.161 |
| | | RO | .233* | .381** | .315** | .221* | .255* | 0.185 |
| | | DO | .323** | 0.194 | .270** | 0.149 | 0.096 | 0.165 |
| | neutral | IG | .265** | .251* | 0.123 | .302** | 0.131 | 0.183 |
| | | RO | .234* | .252* | .255* | .212* | .306** | 0.170 |
| | | DO | 0.154 | .209* | .223* | .208* | .204* | 0.145 |

*Note*. Pearson r – below the diagonal; Spearman Rho – above the diagonal; IG – ingroup; ROG – rival outgroup; NOG neutral outgroup

Table I2

*Experiment 3: Self-unpleasantness ratings - Full correlation table*

| rating | stimuli | group | Self-unpleasantness | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | salient | | | neutral | | |
| | | | IG | RO | DO | IG | RO | DO |
| pain intensity | salient | IG | .637** | .448** | .499** | .396** | .426** | .410** |
| | | RO | .500** | .608** | .472** | .469** | .396** | .499** |
| | | DO | .369** | .412** | .583** | .317** | .269** | .447** |
| | neutral | IG | .404** | .417** | .385** | .597** | .389** | .422** |
| | | RO | .431** | .469** | .415** | .373** | .578** | .447** |
| | | DO | .285** | .476** | .430** | .319** | .270** | .594** |
| self-unpleasantness | salient | IG | 1 | .686** | .753** | .704** | .688** | .714** |
| | | RO | .639** | 1 | .734** | .651** | .733** | .768** |
| | | DO | .747** | .706** | 1 | .700** | .654** | .727** |
| | neutral | IG | .689** | .620** | .676** | 1 | .641** | .679** |
| | | RO | .647** | .765** | .627** | .651** | 1 | .640** |
| | | DO | .697** | .716** | .724** | .649** | .581** | 1 |
| Schadenfreude | salient | IG | -0.118 | 0.032 | 0.057 | 0.047 | 0.093 | 0.012 |
| | | RO | 0.076 | 0.080 | 0.078 | 0.098 | 0.092 | 0.041 |
| | | DO | 0.055 | 0.095 | -0.024 | 0.114 | 0.115 | 0.088 |
| | neutral | IG | 0.064 | 0.180 | 0.105 | -0.086 | 0.178 | 0.136 |
| | | RO | 0.045 | 0.094 | 0.084 | 0.084 | 0.047 | -0.009 |
| | | DO | 0.076 | 0.075 | 0.162 | 0.193 | 0.169 | 0.072 |
| empathic concern | salient | IG | .547** | .486** | .420** | .366** | .436** | .480** |
| | | RO | .326** | .464** | .417** | .409** | .307** | .448** |
| | | DO | .350** | .307** | .497** | .369** | .302** | .391** |
| | neutral | IG | .256* | .367** | .347** | .404** | .380** | .310** |
| | | RO | .249* | .310** | .282** | .318** | .396** | .291** |
| | | DO | .278** | .408** | .389** | .304** | .269** | .465** |
| perspective taking | salient | IG | .271** | .328** | .236* | 0.144 | .379** | .227* |
| | | RO | 0.112 | 0.179 | .201* | 0.167 | 0.169 | 0.172 |
| | | DO | 0.124 | 0.134 | .205* | 0.171 | 0.198 | 0.132 |
| | neutral | IG | 0.083 | .248* | 0.143 | .207* | .266** | 0.098 |
| | | RO | 0.119 | 0.133 | 0.193 | 0.175 | .217* | 0.173 |
| | | DO | 0.025 | 0.121 | 0.152 | 0.103 | 0.171 | 0.116 |

*Note*. Pearson r – below the diagonal; Spearman Rho – above the diagonal; IG – ingroup; ROG – rival outgroup; NOG neutral outgroup

Table I3

*Experiment 3: Schadenfreude ratings - Full correlation table*

| rating | stimuli | group | Schadenfreude | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | salient | | | neutral | | |
| | | | IG | RO | DO | IG | RO | DO |
| pain intensity | salient | IG | -.254* | -0.012 | -0.034 | 0.055 | -0.057 | 0.028 |
| | | RO | -0.134 | -0.193 | 0.049 | -0.023 | -0.089 | 0.015 |
| | | DO | -0.061 | -0.071 | -0.115 | 0.038 | -0.090 | 0.015 |
| | neutral | IG | -0.125 | -0.015 | -0.008 | -0.117 | 0.027 | 0.102 |
| | | RO | -0.093 | -0.100 | 0.075 | 0.135 | -.203* | 0.001 |
| | | DO | -0.052 | -0.111 | -0.004 | 0.018 | -0.088 | 0.012 |
| self-unpleasantness | salient | IG | -0.082 | 0.150 | 0.094 | 0.145 | 0.096 | 0.175 |
| | | RO | 0.125 | 0.112 | 0.147 | .220* | 0.115 | 0.191 |
| | | DO | 0.124 | 0.135 | 0.029 | 0.171 | 0.199 | .244* |
| | neutral | IG | 0.098 | 0.152 | 0.106 | -0.026 | .203* | .236* |
| | | RO | 0.144 | 0.084 | 0.134 | .224* | 0.084 | .249* |
| | | DO | 0.109 | 0.116 | 0.130 | 0.147 | 0.110 | 0.169 |
| Schadenfreude | salient | IG | 1 | .577** | .586** | .556** | .702** | .617** |
| | | RO | .498** | 1 | .514** | .520** | .587** | .667** |
| | | DO | .583** | .459** | 1 | .634** | .576** | .556** |
| | neutral | IG | .629** | .430** | .593** | 1 | .466** | .490** |
| | | RO | .565** | .665** | .483** | .451** | 1 | .577** |
| | | DO | .583** | .592** | .522** | .467** | .559** | 1 |
| empathic concern | salient | IG | -0.148 | 0.018 | 0.054 | 0.100 | -0.020 | -0.106 |
| | | RO | -0.102 | -.294** | -0.029 | -0.134 | -0.198 | -0.075 |
| | | DO | 0.122 | 0.010 | 0.096 | 0.120 | 0.026 | 0.192 |
| | neutral | IG | 0.027 | 0.131 | 0.082 | -0.027 | 0.087 | 0.095 |
| | | RO | -0.177 | -.316** | -0.067 | -0.110 | -.253* | -0.133 |
| | | DO | -0.072 | -0.121 | 0.018 | 0.000 | -0.084 | -0.059 |
| perspective taking | salient | IG | -0.091 | -0.137 | 0.004 | 0.096 | -0.110 | -0.187 |
| | | RO | 0.009 | -.245* | -0.020 | -0.076 | -.212* | -0.124 |
| | | DO | 0.162 | -0.084 | 0.066 | 0.050 | -0.036 | 0.107 |
| | neutral | IG | -0.068 | -0.085 | -0.019 | -0.093 | -0.023 | -0.130 |
| | | RO | 0.042 | -.232* | 0.080 | 0.032 | -0.109 | -0.116 |
| | | DO | 0.102 | -0.160 | 0.023 | 0.051 | -0.098 | -0.047 |

*Note*. Pearson r – below the diagonal; Spearman Rho – above the diagonal; IG – ingroup; ROG – rival outgroup; NOG neutral outgroup

Table I4

*Experiment 3: Empathic concern ratings - Full correlation table*

| rating | stimuli | group | Empathic concern | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | salient | | | neutral | | |
| | | | IG | RO | DO | IG | RO | DO |
| pain intensity | salient | IG | .657** | .442** | .532** | .421** | .399** | .395** |
| | | RO | .474** | .735** | .447** | .433** | .521** | .575** |
| | | DO | .407** | .534** | .557** | .315** | .462** | .429** |
| | neutral | IG | .391** | .489** | .400** | .557** | .447** | .408** |
| | | RO | .322** | .370** | .301** | .263** | .567** | .268** |
| | | DO | .400** | .551** | .441** | .388** | .421** | .617** |
| self-unpleasantness | salient | IG | .527** | .331** | .311** | .251* | .272** | .274** |
| | | RO | .485** | .468** | .306** | .357** | .329** | .420** |
| | | DO | .408** | .393** | .456** | .337** | .308** | .363** |
| | neutral | IG | .364** | .411** | .334** | .360** | .320** | .300** |
| | | RO | .412** | .302** | .312** | .350** | .381** | .269** |
| | | DO | .474** | .469** | .341** | .312** | .334** | .461** |
| Schadenfreude | salient | IG | -0.136 | -0.069 | 0.105 | 0.003 | -0.151 | -0.063 |
| | | RO | 0.008 | -.232* | -0.014 | 0.019 | -.300** | -0.153 |
| | | DO | 0.037 | -0.053 | 0.096 | 0.068 | -0.102 | 0.005 |
| | neutral | IG | 0.138 | -0.076 | 0.128 | -0.002 | -0.060 | -0.017 |
| | | RO | -0.044 | -0.118 | 0.071 | 0.060 | -0.188 | -0.049 |
| | | DO | -0.023 | -0.022 | 0.174 | 0.104 | -0.106 | -0.030 |
| empathic concern | salient | IG | 1 | .640** | .601** | .644** | .552** | .633** |
| | | RO | .613** | 1 | .643** | .631** | .701** | .803** |
| | | DO | .612** | .639** | 1 | .604** | .555** | .610** |
| | neutral | IG | .639** | .617** | .587** | 1 | .563** | .686** |
| | | RO | .525** | .687** | .517** | .529** | 1 | .665** |
| | | DO | .628** | .797** | .628** | .684** | .621** | 1 |
| perspective taking | salient | IG | .589** | .361** | .337** | .350** | .482** | .262** |
| | | RO | .332** | .517** | .363** | .354** | .494** | .317** |
| | | DO | .306** | .322** | .504** | .344** | .334** | .259* |
| | neutral | IG | .374** | .391** | .303** | .644** | .388** | .330** |
| | | RO | .373** | .403** | .413** | .420** | .642** | .383** |
| | | DO | .251* | .320** | .316** | .402** | .471** | .303** |

*Note*. Pearson r – below the diagonal; Spearman Rho – above the diagonal; IG – ingroup; ROG – rival outgroup; NOG neutral outgroup

Table I5

*Experiment 3: Perspective taking ratings - Full correlation table*

| rating | stimuli | group | Perspective taking | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | salient | | | neutral | | |
| | | | IG | RO | DO | IG | RO | DO |
| pain intensity | salient | IG | .440** | .257* | .340** | .301** | .252* | 0.173 |
| | | RO | .316** | .407** | .230* | .314** | .295** | .240* |
| | | DO | .346** | .336** | .299** | 0.171 | .291** | .246* |
| | neutral | IG | .225* | .256* | 0.178 | .323** | .247* | .251* |
| | | RO | .356** | .265** | 0.117 | 0.181 | .332** | .240* |
| | | DO | .204* | .219* | 0.189 | .220* | .206* | 0.169 |
| self-unpleasantness | salient | IG | .281** | 0.160 | 0.130 | 0.106 | 0.148 | 0.079 |
| | | RO | .303** | 0.195 | 0.137 | .256* | 0.142 | 0.139 |
| | | DO | .237* | .207* | 0.195 | 0.157 | .209* | 0.170 |
| | neutral | IG | 0.138 | 0.154 | 0.156 | 0.189 | 0.180 | 0.109 |
| | | RO | .346** | 0.170 | .213* | .239* | .220* | 0.187 |
| | | DO | .229* | 0.197 | 0.117 | 0.132 | 0.196 | 0.136 |
| Schadenfreude | salient | IG | -0.048 | 0.006 | 0.155 | -0.021 | 0.035 | 0.075 |
| | | RO | -0.121 | -.227* | -0.040 | -0.163 | -0.194 | -0.172 |
| | | DO | 0.012 | -0.053 | 0.083 | 0.023 | 0.079 | 0.022 |
| | neutral | IG | 0.154 | -0.041 | 0.051 | -0.025 | 0.024 | 0.059 |
| | | RO | -0.100 | -0.161 | 0.031 | -0.019 | -0.033 | -0.034 |
| | | DO | -0.113 | -0.086 | 0.147 | -0.085 | -0.079 | -0.035 |
| empathic concern | salient | IG | .623** | .367** | .317** | .404** | .414** | .282** |
| | | RO | .392** | .529** | .326** | .414** | .433** | .343** |
| | | DO | .409** | .415** | .580** | .362** | .465** | .397** |
| | neutral | IG | .385** | .387** | .371** | .676** | .482** | .458** |
| | | RO | .498** | .506** | .336** | .413** | .668** | .500** |
| | | DO | .299** | .346** | .265** | .360** | .441** | .347** |
| perspective taking | salient | IG | 1 | .662** | .574** | .633** | .682** | .672** |
| | | RO | .677** | 1 | .634** | .653** | .681** | .684** |
| | | DO | .585** | .665** | 1 | .571** | .656** | .594** |
| | neutral | IG | .632** | .659** | .577** | 1 | .643** | .644** |
| | | RO | .689** | .713** | .680** | .632** | 1 | .744** |
| | | DO | .697** | .722** | .627** | .657** | .765** | 1 |

*Note*. Pearson r – below the diagonal; Spearman Rho – above the diagonal; IG – ingroup; ROG – rival outgroup; NOG neutral outgroup

Appendix J

Experiment 3: Full correlation table of Identity, trait empathy and prejudice with empathy ratings

Table J1

*Experiment 3: Empathy ratings and fan identity, trait empathy and prejudice - Full correlation table (Pearson)*

| rating | stimuli | group | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SSIS | BIRGing | CORFing | F | PT | EC | PD | SDO | \|D\| |
| pain intensity | salient | IG | 0.032 | 0.093 | 0.099 | 0.084 | 0.090 | -0.043 | 0.052 | -0.154 | -0.002 |
| | | RO | -0.098 | 0.036 | 0.010 | 0.152 | 0.132 | 0.106 | 0.144 | -.210* | -0.092 |
| | | DO | -0.070 | 0.094 | -0.044 | 0.158 | 0.123 | 0.157 | 0.098 | -.229* | 0.042 |
| | neutral | IG | 0.120 | 0.085 | 0.002 | .223* | 0.132 | 0.202 | .252* | -0.011 | 0.031 |
| | | RO | -0.097 | 0.106 | 0.063 | 0.120 | 0.100 | 0.083 | 0.084 | -.218* | 0.110 |
| | | DO | 0.052 | 0.037 | -0.134 | 0.012 | 0.001 | 0.101 | 0.086 | -0.186 | -0.025 |
| self-unpleasantness | salient | IG | 0.164 | 0.199 | .265** | 0.032 | -0.051 | 0.001 | 0.181 | -0.135 | -0.062 |
| | | RO | 0.057 | 0.136 | 0.195 | 0.003 | 0.082 | 0.034 | 0.135 | -0.157 | -0.001 |
| | | DO | 0.129 | .219* | .221* | -0.007 | -0.036 | 0.092 | .254* | -0.195 | 0.007 |
| | neutral | IG | 0.196 | 0.187 | 0.167 | 0.118 | 0.056 | 0.071 | .334** | -0.079 | 0.040 |
| | | RO | 0.037 | 0.092 | .206* | 0.044 | 0.105 | 0.103 | 0.165 | -0.179 | -0.075 |
| | | DO | 0.133 | 0.109 | 0.173 | -0.056 | -0.048 | 0.062 | .208* | -.223* | -0.057 |
| Schadenfreude | salient | IG | 0.197 | .281** | 0.043 | -0.159 | -0.038 | -0.088 | 0.104 | 0.182 | -0.022 |
| | | RO | .202* | .263** | 0.017 | -0.055 | 0.014 | -0.109 | 0.145 | .258* | 0.191 |
| | | DO | .238* | .322** | 0.101 | -0.009 | 0.056 | -0.139 | 0.098 | 0.083 | 0.119 |
| | neutral | IG | 0.191 | .276** | 0.103 | -0.111 | -0.075 | -0.184 | -0.035 | 0.125 | 0.041 |
| | | RO | .274** | .270** | 0.138 | -0.014 | -0.045 | -0.039 | 0.183 | .259* | 0.173 |
| | | DO | .234* | .249* | -0.086 | -0.041 | -0.055 | 0.087 | 0.146 | .224* | 0.000 |
| empathic concern | salient | IG | -0.003 | 0.053 | 0.055 | 0.140 | .250* | 0.034 | -0.020 | -.289** | -0.034 |
| | | RO | -0.164 | -0.052 | 0.047 | 0.105 | .238* | .227* | 0.010 | -.374** | -0.201 |
| | | DO | 0.103 | .206* | 0.085 | 0.138 | 0.161 | 0.192 | 0.170 | -.309** | -0.179 |
| | neutral | IG | 0.020 | -0.018 | 0.082 | .281** | .396** | .236* | 0.090 | -0.171 | -0.133 |
| | | RO | -0.116 | -0.108 | 0.042 | .260* | .353** | .316** | -0.067 | -.428** | -0.099 |
| | | DO | 0.027 | -0.037 | 0.034 | 0.069 | 0.169 | 0.177 | 0.031 | -.284** | -0.136 |
| perspective taking | salient | IG | -.300** | -0.122 | 0.117 | .300** | .212* | 0.094 | -0.141 | -.370** | -0.102 |
| | | RO | -.366** | -0.163 | .253* | 0.166 | 0.162 | 0.084 | -0.144 | -.445** | -0.167 |
| | | DO | -0.156 | -0.011 | 0.106 | 0.121 | 0.188 | 0.106 | -0.027 | -.242* | -0.160 |
| | neutral | IG | -.233* | -0.126 | 0.176 | .287** | .358** | 0.067 | -0.059 | -.285** | -0.162 |
| | | RO | -.277** | -0.138 | .211* | .297** | .268** | 0.165 | -0.038 | -.343** | -0.154 |
| | | DO | -0.191 | -0.070 | .210* | .322** | 0.149 | 0.172 | -0.134 | -.294** | -0.199 |

Note. IG – ingroup; ROG – rival outgroup; NOG neutral outgroup; SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; |D| – Implicit associations test D measure, absolute value; * $p < .05$; ** $p < .01$; N = 97

Table J2

*Experiment 3: Empathy ratings and fan identity, trait empathy and prejudice - Full correlation table (Spearman)*

| rating | stimuli | group | Fan identity | | | Interpersonal reactivity index | | | | Prejudice | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SSIS | BIRGing | CORFing | F | PT | EC | PD | SDO | \|D\| |
| pain intensity | salient | IG | -0.016 | 0.031 | 0.087 | 0.096 | 0.147 | -0.060 | 0.033 | -0.148 | -0.016 |
| | | RO | -0.108 | 0.015 | 0.011 | 0.187 | 0.141 | 0.073 | 0.147 | -.270** | -0.134 |
| | | DO | -0.106 | 0.050 | -0.043 | 0.186 | 0.161 | 0.150 | 0.106 | -.232* | 0.014 |
| | neutral | IG | 0.083 | 0.040 | 0.012 | .239* | 0.114 | 0.166 | .228* | -0.045 | -0.016 |
| | | RO | -0.082 | 0.089 | 0.067 | 0.073 | 0.089 | 0.046 | 0.121 | -0.198 | 0.082 |
| | | DO | 0.035 | 0.045 | -0.117 | 0.052 | 0.051 | 0.078 | 0.081 | -0.183 | 0.011 |
| self-unpleasantness | salient | IG | 0.157 | 0.179 | .282** | 0.023 | -0.026 | -0.044 | 0.147 | -0.099 | -0.119 |
| | | RO | 0.123 | 0.168 | .213* | -0.005 | 0.118 | 0.012 | 0.153 | -0.145 | -0.024 |
| | | DO | 0.163 | .249* | .261** | 0.021 | 0.030 | 0.082 | .258* | -0.172 | -0.041 |
| | neutral | IG | .219* | 0.160 | .212* | 0.097 | 0.041 | 0.060 | .328** | -0.094 | -0.021 |
| | | RO | 0.092 | 0.093 | .217* | -0.007 | 0.082 | 0.088 | 0.179 | -0.147 | -0.109 |
| | | DO | 0.146 | 0.140 | .213* | -0.067 | -0.016 | 0.039 | .241* | -.209* | -0.055 |
| Schadenfreude | salient | IG | 0.174 | 0.192 | -0.020 | -.249* | 0.015 | -0.083 | 0.099 | 0.182 | -0.007 |
| | | RO | 0.187 | .267** | -0.020 | -0.143 | -0.061 | -0.154 | 0.157 | .325** | 0.167 |
| | | DO | .200* | .285** | 0.110 | -0.098 | -0.017 | -0.153 | 0.163 | 0.130 | 0.069 |
| | neutral | IG | 0.146 | .256* | 0.098 | -0.132 | -0.015 | -.207* | -0.024 | 0.149 | 0.019 |
| | | RO | .294** | .261** | 0.104 | -0.041 | -0.069 | -0.021 | .222* | .211* | 0.083 |
| | | DO | .230* | .263** | -0.103 | -0.072 | -0.065 | 0.076 | 0.140 | 0.200 | 0.047 |
| empathic concern | salient | IG | -0.010 | 0.017 | 0.058 | 0.179 | .286** | 0.036 | -0.033 | -.271** | -0.058 |
| | | RO | -0.181 | -0.062 | 0.060 | 0.134 | .249* | 0.198 | 0.004 | -.368** | -.217* |
| | | DO | 0.050 | 0.187 | 0.104 | 0.150 | 0.176 | 0.202 | 0.135 | -.289** | -0.190 |
| | neutral | IG | -0.005 | -0.006 | 0.107 | .333** | .401** | .242* | 0.066 | -0.160 | -0.121 |
| | | RO | -0.145 | -0.118 | 0.038 | .236* | .355** | .295** | -0.052 | -.439** | -0.128 |
| | | DO | -0.005 | -0.024 | 0.043 | 0.121 | 0.179 | 0.160 | 0.031 | -.271** | -0.143 |
| perspective taking | salient | IG | -.258* | -0.158 | 0.087 | .317** | .248* | 0.098 | -0.148 | -.369** | -0.102 |
| | | RO | -.338** | -0.155 | .230* | 0.193 | 0.151 | 0.073 | -0.128 | -.400** | -0.159 |
| | | DO | -0.120 | 0.004 | 0.098 | 0.154 | 0.169 | 0.118 | -0.016 | -.253* | -0.164 |
| | neutral | IG | -.217* | -0.120 | 0.167 | .322** | .360** | 0.070 | -0.088 | -.253* | -0.149 |
| | | RO | -.252* | -0.127 | 0.179 | .315** | .269** | 0.156 | -0.028 | -.337** | -0.139 |
| | | DO | -0.149 | -0.052 | 0.175 | .332** | 0.129 | 0.169 | -0.120 | -.283** | -0.183 |

Note. IG – ingroup; ROG – rival outgroup; NOG neutral outgroup; SSIS – Sport spectator identification scale, F – Fantasy, PT – Perspective taking, EC – Empathic concern, PD – Personal distress, SDO – Social dominance orientation; |D| – Implicit associations test D measure, absolute value; * $p < .05$; ** $p < .01$; N = 97

Appendix K

Experiment 4: Accuracy

Table K1

*Percentage of accurate and inaccurate pain victim identification per stimulus*

| perpetrator - victim | Stimulus number | % accurate | painfulness | perpetrator - victim | Stimulus number | % accurate | painfulness |
|---|---|---|---|---|---|---|---|
| Partizan – Red Star | 1 | 92.4% | 3.29 | Other club – Red Star | 1 | 97.5% | 3.73 |
| | 2 | 94.9% | 2.77 | | 2 | 96.2% | 2.39 |
| | 3 | 95.6% | 4.09 | | 3 | 58.2% | 3.02 |
| | 4 | 92.4% | 3.11 | | 4 | 97.5% | 3.13 |
| | 5 | 98.7% | 3.30 | | 5 | 97.5% | 2.88 |
| | 6 | 90.5% | 3.73 | | 6 | 97.5% | 4.42 |
| | 7 | 97.5% | 3.05 | | 7 | 98.1% | 4.14 |
| | 8 | 98.1% | 4.05 | | 8 | 96.8% | 3.12 |
| | 9 | 90.5% | 2.65 | | 9 | 98.1% | 2.69 |
| | 10 | 98.1% | 2.93 | | 10 | 86.7% | 3.64 |
| Red Star – Partizan | 1 | 99.4% | 2.94 | Other club - Partizan | 1 | 99.4% | 4.86 |
| | 2 | 75.3% | 3.13 | | 2 | 98.7% | 4.46 |
| | 3 | 93.7% | 4.87 | | 3 | 98.7% | 2.95 |
| | 4 | 81.6% | 3.04 | | 4 | 93.7% | 3.51 |
| | 5 | 95.6% | 2.95 | | 5 | 95.6% | 3.91 |
| | 6 | 98.7% | 3.24 | | 6 | 72.8% | 2.97 |
| | 7 | 98.1% | 4.46 | | 7 | 97.5% | 3.18 |
| | 8 | 90.5% | 3.26 | | 8 | 99.4% | 3.11 |
| | 9 | 66.5% | 3.67 | | 9 | 98.1% | 3.97 |
| | 10 | 91.8% | 3.12 | | 10 | 98.7% | 3.90 |
| Red Star – Other club | 1 | 98.7% | 4.42 | Partizan – Other club | 1 | 93.0% | 2.96 |
| | 2 | 78.5% | 2.93 | | 2 | 94.9% | 3.05 |
| | 3 | 98.1% | 4.50 | | 3 | 93.7% | 3.41 |
| | 4 | 96.8% | 3.88 | | 4 | 98.1% | 2.10 |
| | 5 | 65.8% | 5.02 | | 5 | 96.8% | 4.77 |
| | 6 | 59.5% | 3.53 | | 6 | 98.1% | 2.48 |
| | 7 | 97.5% | 2.78 | | 7 | 97.5% | 2.58 |
| | 8 | 61.4% | 3.34 | | 8 | 92.4% | 2.73 |
| | 9 | 68.4% | 2.73 | | 9 | 90.5% | 3.52 |
| | 10 | 93.0% | 2.44 | | 10 | 75.9% | 3.66 |

*Note*. Excluded stimuli are marked in red

Appendix L

Table L1

*Post hoc pairwise comparisons (Bonferroni corrected) for different perpetrator – fouled player combinations*

| Perpetrator_victim | | Mean Difference | Sig. | Perpetrator_victim | | Mean Difference | Sig. |
|---|---|---|---|---|---|---|---|
| par_cz | dr_cz | -0.040 | 1.000 | dr_par | par_cz | .419* | 0.000 |
| | cz_par | -0.183 | 0.552 | | dr_cz | .380* | 0.000 |
| | dr_par | -.419* | 0.000 | | cz_par | .236* | 0.000 |
| | cz_dr | -.267* | 0.019 | | cz_dr | .152* | 0.020 |
| | par_dr | .184* | 0.001 | | par_dr | .604* | 0.000 |
| dr_cz | par_cz | 0.040 | 1.000 | cz_dr | par_cz | .267* | 0.019 |
| | cz_par | -0.143 | 1.000 | | dr_cz | .227* | 0.032 |
| | dr_par | -.380* | 0.000 | | cz_par | 0.084 | 1.000 |
| | cz_dr | -.227* | 0.032 | | dr_par | -.152* | 0.020 |
| | par_dr | .224* | 0.000 | | par_dr | .451* | 0.000 |
| cz_par | par_cz | 0.183 | 0.552 | par_dr | par_cz | -.184* | 0.001 |
| | dr_cz | 0.143 | 1.000 | | dr_cz | -.224* | 0.000 |
| | dr_par | -.236* | 0.000 | | cz_par | -.367* | 0.000 |
| | cz_dr | -0.084 | 1.000 | | dr_par | -.604* | 0.000 |
| | par_dr | .367* | 0.000 | | cz_dr | -.451* | 0.000 |

*Note*. sig – significance; par – Partizan; cz – Red Star; dr – Other club

## Biography

Marija Čolić (1991, Belgrade) completed Milan Rakić elementary school in Gornja Toplica in 2006, and Valjevo gymnasium in 2010. She enrolled in Psychology studies at the University of Belgrade – Faculty of Philosophy the same year, where she graduated with a bachelor's degree in 2014 (GPA 9.75/10) and a master's degree (GPA 10/10) in 2016. She defended her master's thesis, *The evaluative role of visual perspective in autobiographical memories: Why do we recall some events from the 1st person and other from the 3rd person?*, under the supervision of Iris Žeželj, who continued to be her supervisor during her PhD studies at the same University. She received several awards and stipends during her undergraduate and postgraduate education.

Marija is employed as a teaching assistant at the University of Belgrade – Faculty of Sports and Physical Education, teaching General, Developmental, and Educational Psychology courses. She was previously employed at the Institute for Medical Research, in the Neurophysiology research group. Before that, she was engaged in assisting with undergraduate courses (Psychology of Individual Differences, Psychometrics, and Principles of Psychological Testing) at her Faculty.

Marija published 11 papers in domestic and international journals and participated in many scientific conferences. During her studies, she participated in two reproducibility projects and several other international scientific collaborations.

# Изјава о ауторству

Име и презиме аутора: *Марија Чолић*
Број индекса: *4П16-1*

**Изјављујем**

да је докторска дисертација под насловом

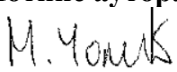
*Intergroup Empathy – Conceptual Systematisation and Empirical Contribution to Understanding the Phenomenon*

• резултат сопственог истраживачког рада;
• да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
• да су резултати коректно наведени и
• да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

У Београду,

*01.09.2023.*

**Потпис аутора**



# Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: ***Марија В. Чолић***

Број индекса: ***416П-1***

Студијски програм: ***Психологија***

Наслов рада: ***Intergroup Empathy – Conceptual Systematisation and Empirical Contribution to Understanding the Phenomenon***

Ментор: ***др Ирис Л. Жежељ***

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

У Београду,

*01.09.2023.*

**Потпис аутора**

М. Чолић

# Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

***Intergroup Empathy – Conceptual Systematisation and Empirical Contribution to Understanding the Phenomenon***

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

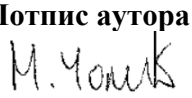
Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
**2. Ауторство – некомерцијално (CC BY-NC)**
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)
(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

У Београду,

*01.09.2023.*

**Потпис аутора**



1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.