HELPING "US" VS. "THEM": INGROUP FAVOURITISM IN PROSOCIAL BEHAVIOUR



Jim Albert Charlton Everett

Corpus Christi College

Thesis Submitted for the Degree of

Doctor of Philosophy

of

The University of Oxford

Preface

Abstract

To what extent is the tendency to act more prosocially towards ingroup than outgroup members a 'default' tendency that is based on intuitive processes activated across different social contexts and with different social groups? Here I report three sets of targeted empirical studies that use economic games to explore ingroup favouritism in prosocial behaviour. In Chapter 2, I use a newly developed measure to explore whether simply categorizing people into groups is sufficient to induce ingroup favouritism in both gains and losses frames. I find that minimal groups are significantly more prosocial to ingroup members than outgroup members, regardless of the economic cost of helping, whether decisions were public or private, and whether the decision is framed in terms of gains and losses. In Chapter 3 I apply dual-process models to intergroup prosocial behaviour and show that even under time pressure without the potential to deliberate, participants cooperate more with minimal ingroup than outgroup members in both gains and losses frames. In Chapter 4, I explore ingroup favouritism in the context of real religious groups using games that either have (Trust Game) or lack (Dictator Game) interdependent outcomes. I find ingroup favouritism only by atheist participants in a Dictator Game. In both games, Christian participants transferred more to both Christian and atheist recipients, and this was driven by the frequency of thinking about religious ideas. Together, my results support the thesis that ingroup favouritism is a default, intuitive, tendency robust across different contexts and even with novel groups, and charts a course for future research.

Table of Contents

PREFACE **II**

ABSTRACT III

ACKNOWLEDGEMENTS VIII

COLLABORATIVE WORK AND DECLARATION X

LIST OF ABBREVIATIONS XIII

LIST OF TABLES XIII

LIST OF FIGURES XIII

CHAPTER 1: INTRODUCTION **1**

BACKGROUND: "US", "THEM", AND PROSOCIAL BEHAVIOUR **3**

INTERGROUP RELATIONS 3

PROSOCIAL BEHAVIOUR 11

GAINS AND LOSSES IN INTERPERSONAL DECISION MAKING 23

EXPLAINING INGROUP FAVOURITISM IN PROSOCIAL BEHAVIOUR **35**

SOCIAL IDENTITY AND 'MERE CATEGORIZATION' 37

RECIPROCITY AND REPUTATION 52

OVERVIEW OF CHAPTERS **64**

CHAPTER 2: MERE CATEGORIZATION IN GAINS AND LOSSES CONTEXTS 64

CHAPTER 3: INTUITIVE INGROUP FAVOURITISM 66

CHAPTER 4: RECIPROCITY AND RELIGIOUS PROSOCIALITY 67

CHAPTER 2: MERE CATEGORIZATION IN GAINS AND LOSSES FRAMES. **70**

ABSTRACT **71**

INTRODUCTION **71**

Cost of Helping, Economic Efficiency, and Utility Calculations 74

Within-Subjects Analysis 81

Gains and Losses 82

Present Research **84**

Method 86

Results 94

Discussion **101**

Contributions 101

Limitations 104

Conclusion 106

CHAPTER 3: INTUITIVE INGROUP FAVOURITISM **107**

Abstract **108**

Introduction **108**

Intuitive Prosociality 110

Intuitive Prosociality in Intergroup Contexts 118

Intuitive Prosociality in Gains and Losses Contexts 125

Present Research **128**

Method 128

Results 132

Discussion **135**

Contributions 136

Limitations 141

Conclusion 142

CHAPTER 4: RECIPROCITY AND RELIGIOUS PROSOCIALITY **144**

ABSTRACT **145**

INTRODUCTION **145**

EXPECTATIONS OF RECIPROCITY IN INGROUP FAVOURITISM 146

RELIGIOUS PROSOCIALITY 149

STUDY 1 **154**

METHOD 154

RESULTS 159

STUDY 2 **163**

METHOD 163

RESULTS 165

DISCUSSION **168**

CONTRIBUTIONS 169

LIMITATIONS 171

CONCLUSION 175

CHAPTER 5: GENERAL DISCUSSION **176**

SUMMARY OF RESEARCH FINDINGS **177**

CONTRIBUTIONS **179**

INGROUP FAVOURITISM IN GAINS AND LOSSES 180

'MERE CATEGORIZATION' VS. SOCIAL EXCHANGE 182

INTUITIVE INGROUP FAVOURITISM 183

LIMITATIONS **184**

MTURK 184

DECEPTION 194

DIRECTIONS FOR FUTURE RESEARCH **197**

INTUITIVE INTERGROUP AGGRESSION 197

PAROCHIAL MORALITY 200

CONCLUDING REMARKS **204**

REFERENCES **206**

APPENDIX **233**

STUDY 1 (DICTATOR GAME) 233

STUDY 2 (TRUST GAME) 238

Acknowledgements

First, the Economic and Social Research Council, the Fulbright Commission, and the various other organizations who have supported my attendance at summer schools and conferences. Your funding made my work possible, and I am forever grateful. As Liza Minelli sang in *Cabaret*, money makes the world go 'round, and thanks to your money I've been able to go around the world, conduct my research, and learn from some of the best minds in the field.

Second, the various people who have supervised me. I cannot thank you enough. Miles Hewstone, for starting me on this journey and inculcating a passion for the study of intergroup relations. David Rand, for your infectious enthusiasm in inspiring me to look deeper into the economic and evolutionary roots of social behaviour. Fiery Cushman, for continually being two mental steps ahead of me but always taking the time and effort to let me catch up. Mina Cikara, for challenging and pushing me to become the best scholar I can. Molly Crockett and Nadira Faber: You have both been the most wonderful mentors I could have ever have dreamed of. You never stopped supporting me and have kept faith in me when I was losing it myself, and have continually pushed me onwards and upwards in my academic trajectory. I will never be able to thank you enough. I just hope that one day I can be even half the supervisor to someone else that you have been to me.

Third, the various colleagues, collaborators, and mentors who have taught me so much. You have inspired, instructed, and invigorated me in so many ways, even when you might not have realized you were doing it.

Fourth, the many, many friends I have made during this journey. I have known you from childhood, I have met you at conferences, I have spent time with you at summer schools, and I have been welcomed into your labs. You have made me laugh, you have introduced me to your countries and customs, and you have all taught me so much. You all rock.

Fifth, Robert and Garrett, for being my family away from my family, and a home away from home. All those years ago when I was still in high school, you convinced me that I should be applying to university, and then convinced me that I could apply to Oxford. Since then, you have supported me and given me my much-needed education into gay history and culture. And perhaps most importantly, you introduced me to *Ru Paul's Drag Race*, thus giving me a sickening back-up plan if this academic gig doesn't work out.

Sixth, my (biological) family who have tirelessly supported me, both financially and emotionally, through this long period of academic training. I – literally – wouldn't be here without you, and I am forever grateful. Maybe one day I'll repay my thanks and get a "real job". Hopefully not, though.

And finally, thank you to Maggie, my little Iron Lady. You've kept me sane(ish) in the most insane of times.

Collaborative work and declaration

For all this work, I am grateful to my collaborators who provided feedback and suggestions throughout the process. Across all the studies, I had primary responsibility for the design, data collection, analysis, and write up. I provide details of my exact contribution below, which all collaborators have approved.

Chapter 2: Mere Categorization in Gains and Losses Contexts

Nadira Faber (University of Oxford)

Molly Crockett (University of Oxford)

Everett, Faber, and Crockett developed the study concept. Crockett had previously created the initial lottery task, which Everett adapted for an intergroup context. Everett, Crockett, and Faber participated in finalizing the design. Everett coordinated the study and collected data, and conducted analysis with feedback from Faber and Crockett. Everett drafted the manuscript with critical revisions from Faber and Crockett.

Everett, J. A. C., Faber, N. S., & Crockett, M. J. (2015). The influence of social preferences and reputational concerns on intergroup prosocial behaviour in gains and losses contexts. *Royal Society Open Science*, *2*(12), 150546. https://doi.org/10.1098/rsos.150546

Chapter 3: Intuitive Ingroup Favouritism

Mina Cikara (Harvard University)

Zach Ingbretsen (Harvard University)

Fiery Cushman (Harvard University)

Everett, Cikara, and Cushman developed the study concept. Everett and Ingbretsen implemented the study online and collected the data. Everett and Ingbretsen conducted data analysis, with input from both the other authors. Everett drafted the manuscript with critical revisions from Cikara and Cushman.

Everett, J.A.C., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behavior–even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *Journal of Experimental Social Psychology*. 73, 76-81. https://doi.org/10.1016/j.jesp.2017.06.014

Chapter 4: Reciprocity and Religious Prosociality

David Rand (Yale University)

Omar Sultan Haque (Harvard University)

Everett, Haque, and Rand developed the study concept. Everett and Rand implemented the study online and collected the data. Everett conducted data analysis with input from Rand. Everett drafted the manuscript with critical revisions from Haque and Rand.

Everett, J. A. C., Haque, O. S., & Rand, D. G. (2016). How good is the Samaritan, and why? An experimental investigation of the extent and nature of religious prosociality using economic games. *Social Psychological and Personality Science*, *7*(3), 248–255. https://doi.org/10.1177/1948550616632577

List of Abbreviations

BGR	Bounded Generalised Reciprocity
CPD	Common Pool Dilemma
CRS	Centrality of Religion Scale
DG	Dictator Game
EV	Expected Value
ILT	Intergroup Lottery Task
PGD	Public Goods Dilemma
SIT	Social Identity Theory
SVO	Social Values Orientation
TG	Trust Game

List of Tables

Table 2.1: Helping as a function of condition. 97

Table 3.1. Rates of cooperation across conditions. 135

List of Figures

Figure 1.1: The Prisoner's Dilemma 19

Figure 1.3: The Trust Game 22

Figure 1.4: The Tajfel Matrices 38

Figure 2.1: Example decision screen for the gains-context lottery task (Study 1). 90

Figure 2.2: Example decision screen for the losses-context lottery task (Study 2) 91

Figure 2.3: Helping as a function of group and decision context in Study 1 96

Figure 2.4: Helping as a function of group and decision context in Study 2 96

Figure 2.5: Helping as a function of group and economic efficiency in Study 1 99

Figure 2.6: Helping as a function of group and economic efficiency in Study 2 99

Figure 3.1. Example Matrices given to participants (left: gains frame with in-group partner; right: losses frame with outgroup partner) 131

Figure 3.2: Cooperation in the PD as a function of gains/losses frame, group membership of the other player, and whether decisions were made under time pressure, time delay, or no timing instructions. 134

Figure 4.1: Transfers in cents as a function of self-identification as Christian or Atheist and target religiosity. Error bars represent 95% C.I.s 160

Figure 4.2: Transfers in cents as a function of recipient and participant religiosity (binned to the nearest integer with the mean displayed across all observations in that bin). The size of the dots is proportional to the number of observations. 161

Figure 4.3: Transfers in cents as a function of self-identification as Christian or Atheist and target religiosity. Error bars represent 95% C.I.s 166

Figure 4.4: Transfers in cents as a function of recipient and participant religiosity (binned to the nearest integer with the mean displayed across all observations in that bin). The size of the dots is proportional to the number of observations. 167

Chapter 1: Introduction

From the dawn of our species to the present day, humans have lived and survived in groups. These groups have expanded from small, primarily kin-based ties to groups based on language, religion, geographical location, and even seemingly arbitrary characteristics such as the ownership of a specific brand of electronic device. As a species, we appear to have a remarkable tendency to seek out and identify with people like "us" while distancing ourselves from "them". Indeed, it is in our group-based character that the angels and demons of human nature can be seen: on the one hand, the shining success of intragroup cooperation that has given us democracy and civil rights; and on the other hand, the darkness of intergroup conflict that has given us the collective stains on human history of genocide and war.

In this thesis, I explore the phenomenon of *prosocial behaviour*: providing a benefit to another at a cost to the self. In particular, I explore prosocial behaviour in intergroup contexts and look at the well-documented tendency for individuals to act more prosocially towards ingroup than outgroup members: what I will call *ingroup-favouring prosocial behaviour*, or *ingroup favouritism* for short. Using a methodology of economic games where prosociality is operationalized as giving up some money to benefit another, I look at when and why individuals are more prosocial towards ingroup than outgroup members. I report three targeted empirical chapters that each adopt a different perspective to look at prosocial behaviour extended towards ingroup and outgroup members, and in each chapter make distinct contributions to understanding both prosocial behaviour generally, and ingroup-favouring prosocial behaviour specifically. Together, my results paint a picture in which acting more prosocially towards ingroup

than outgroup members is a 'default' position based on intuitive process that are activated across different contexts and even with novel groups.

Background: "Us", "Them", and Prosocial Behaviour

One of the most enduring findings in social psychology is the tendency for *intergroup bias*: "the systematic tendency to evaluate one's own membership group (the in-group) or its members more favorably than a nonmembership group (the out-group) or its members" (Hewstone, Rubin, & Willis, 2002, p. 576). Intergroup bias manifests in a variety of cognitive (e.g. stereotyping), affective (e.g. prejudice), and behavioural (e.g. discrimination) ways, and the manifestation that I am interested in here is the extension of prosocial behaviour. In this section I shall provide a very brief overview of some of the necessary background on intergroup relations and prosocial behaviour to lay the groundwork for the empirical work reported in this thesis.

Intergroup Relations

What is a Social Group?

Within social psychology, an influential approach to understanding group processes is the social identity perspective, based on the sister-theories of *Social Identity Theory* (SIT: Tajfel & Turner, 1979) and *Self-Categorization Theory* (SCT: Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). According to this approach, a social group is a collection of "two or more individuals who share a common social identification of themselves or, which is nearly the same thing, perceive themselves to be members of the same social category" (Turner, 1982, p. 15). This critical focus here is on *social identity*:

"that part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership" (Tajfel, 1974, p. 69).

Other scholars have, however, disputed this definition. Specifically, Rabbie and colleagues (1969, 1989) have argued against this social identity approach and contested that mere social categorization or social identification is enough to create a social group. Instead, they argue, social categories do not become social groups until perceived interdependence of outcomes fosters a belief that members of different categories will act for or against an individual's self interest (Rabbie & Horwitz, 1969; Rabbie, Schot, & Visser, 1989). According to this perspective, it is important to distinguish between a social category and a social group as explanatory concepts (Rabbie & Horwitz, 1988). Specifically, a social category is "a collection of individuals who share at least one attribute in common", while "a social group can be considered as a 'dynamic whole' or social system, characterized by the perceived interdependence among its members" (Rabbie & Horwitz, 1988, p. 117). In other words, one cannot assume[1] that just because an individual shares an attribute in common with another, that they therefore share a

[1] That said, despite the claims of Rabbie and colleagues, it is not entirely clear that SIT and SCT fail to make this distinction. Turner is clear that "the group is both a psychological process and a social product" (Turner & Giles, 1981, p. 27) but that the social identity perspective is concerned with group membership as a "psychological state, and not a formal-institutional state, with the subjective sense of togetherness, we-ness, or belonging-ness which indicates the formation of a psychological group" (Turner, 1982, p. 16). In an attempt to refute what he sees as a mischaracterization of social groups and categories in the social identity approach, Turner is explicit: "a group not a category, but a group may emerge from the social and psychological processes which lead a collection of people to invent, create, and use a social category" (Turner & Bourhis, 1996, p. 33). It seems, then, that despite back-and-forth, Rabbie and colleagues are actually in agreement with social identity theorists when they claim that there is an important difference between social groups and social categories

social identity. Recognising these points, I shall adopt the following definition of a group in my thesis: a social group is a collection of individuals who perceive themselves (i.e. have a social identity) as members of a shared social category that offers some degree of common fate and interdependence of outcomes. Note, however, that while groups in this social psychological sense are domain general such that the same kind of motivations and social identity processes drive identification with all kinds of groups, this assumption is challenged by scholars working within a more evolutionary framework (Cosmides et al., 2003; Kurzban, Tooby, & Cosmides, 2001; Pietraszewski, Cosmides, & Tooby, 2014). That said, it is unfortunately beyond the scope of this thesis to go deeper into this debate, and so I shall focus on the definition of groups as put forward and used by the researchers and social-psychological theories that I draw most heavily on in this thesis.

Why do we Form Social Groups?

There are a number of different social psychological theories that focus on why individuals form groups. Before delving into these, however, it is useful to briefly note some of the more evolutionary-based explanations of why we form social groups. A powerful argument at the ultimate level for why we form groups comes from evolutionary theory and posits that groups – or coalitions - provided material resource benefits and so, over time, enhanced survival (Brewer, 2004; Henrich & Henrich, 2007; Richerson & Boyd, 2005). The prehistoric environment in which humans evolved was an extremely harsh one and a lone human would, especially compared to some of our hardier animal cousins, face a life that was "nasty, brutish, and short" (Hobbes, 1668/1994). Humans living in small groups, however, could cooperate through sharing

tasks and aiding each other in times of need, thus enhancing each person's survival (Alvard, 2001; Connor, 1995; Harcourt & de Waal, 1992). Such a tendency to cooperate, however, comes with an important caveat: it must be reciprocated. Indiscriminately trusting and cooperating with non-relatives who will not respond in turn would be a distinctly ineffective survival strategy. Rather, it was more effective to cooperative selectively within smaller social groups for whom trust that such behaviour would be reciprocated was higher (Brewer, 2004; Caporael, 1997; Tooby & Cosmides, 2010). Because those individuals who could live effectively in groups would be more likely to survive and pass on their genes, at an ultimate level evolution has selected for a tendency to seek out, identify with, and be oriented towards groups (Alvard, 2001; Brewer, 2004; Caporael, 1997; Connor, 1995; Cosmides, Tooby, & Kurzban, 2003; Harcourt & de Waal, 1992; Kurzban & Leary, 2001; Tooby & Cosmides, 2010). While interesting and important as deep background to the work of this thesis, in my work I will focus on the proximate, social-psychological level.

Traditionally (at least before the advent of SIT), social psychologists took a functional approach to looking at why individuals form groups. Specifically, *realistic group conflict theory* (RCT) (Campbell, 1965; LeVine & Campbell, 1972; Sherif, Harvey, White, Hood, & Sherif, 1961) focuses on the functional relationships between groups of people to suggest that "real conflict of group interests causes intergroup conflict" (Campbell, 1965, p. 287). Like theories posited more at the ultimate level, this account suggests that people form groups and engage in intergroup conflict to better obtain rewards. The advantage of such a functional, resource-based account is that it plays close

attention to the material consequences of group living, and thus provides an account of how forming groups is psychologically rational. A central problem with this approach (at the proximate level, at least) is that does a better job explaining the psychological consequences of group formation rather than the motivational processes that underlie the development and maintenance of group identities.

A number of motivational accounts of group formation have been offered in social psychology, including SIT and SCT (Tajfel & Turner, 1979; Turner et al., 1987), optimal distinctiveness theory (Brewer, 1991, 2011), subjective uncertainty reduction theory (Hogg, 2000; Hogg & Abrams, 1993), and terror management theory (Solomon, Greenberg, & Pyszczynski, 1991). I shall focus here on SIT and SCT as the 'original' motivational account and the one that is most relevant to the work reported in this thesis, though I direct readers to excellent existing overviews on the relationships between these theories (e.g. Hewstone et al., 2002).

SIT is an overarching theory that has been discussed in almost all areas of social psychology, but it is also one that has undergone many expansions, refinements, and updates that mean it can be difficult to pin down exactly what the theory says. It is useful, then, to integrate and summarise the key tenets and central predictions of 'classic' SIT as systematized in Tajfel and Turner (1979), and the later development of SCT as described by (Turner et al., 1987). In doing this, I shall favour conciseness and strict accuracy by quoting directly from Ellemers and Haslam (2011), who identify three core assumptions of SIT:

> "(1) To the extent that individuals internalize a group membership as a meaningful aspect of their self-concept, they will strive to make favorable comparisons between this group and relevant outgroups, in order to achieve or maintain a positive social identity. (2) As a result, social categorization can be sufficient to engender intergroup discrimination and inter-group conflict (i.e., in the absence of a conflict of interest over the division of resources or material outcomes, for example, as a result of historical antagonism). (3) The search for positive social identity may take different forms (individual mobility, social creativity, social competition), depending on consensual definitions of social reality that pertain to socially shared justifications (legitimacy of group and individual outcomes) and perceived cognitive alternatives to current status relations (permeability of group boundaries and stability of status relations)."
>
> (Ellemers & Haslam, 2011, p. 384)

One decade later, Turner and his colleagues outlined self-categorization theory: a sister-theory to SIT that focused more on the cognitive processes underpinning group behaviour. In contrast to SIT, the later SCT places more emphasis on the fact that groups exist at different levels of inclusiveness or abstraction and that people belong to a number of social groups existing at different levels of inclusiveness or abstraction (e.g. 'English', 'British', 'European'). In line with other scholars (e.g. Balliet, Wu, & De Dreu, 2014), I shall refer to both SIT and SCT under the heading of SIT since they are deeply related and for my purposes, offer equivalent predictions.

How does SIT explain why we form groups? According to SIT, external features of the environment activate domain-general processes of categorization, causing individuals to self-categorize by interpreting the social world as consisting of *ingroups*— social groups to which the individual belongs—and *outgroups*— social groups of which the individual is not a member. More specifically, *depersonalization* refers to the psychological process through which people come to perceive the self as an interchangeable exemplar of a social category, rather than a separate individual with unique traits (Tajfel & Turner, 1979; Turner et al., 1987). When individuals categorize themselves as group members, the ingroup becomes integrated with the self and individuals come to recognize the characteristics of the ingroup as representing part of themselves (E. R. Smith, Coats, & Walling, 1999; Tropp & Wright, 2001). Moreover, because the stereotypes and perceptions of the groups are typically ingroup favouring, when placed in an intergroup context individuals achieve positive self-esteem and self-enhancement by focusing on their social identity. Put simply, according to SIT people form groups because (external) social categorization influences the way that people categorize themselves, which in turn influences their self-concept.

How do Social Groups Influence Behaviour?

Social groups are ubiquitous and so are the ways they influence behaviour. Identifying and acting as a member a social group has such a range of cognitive, emotional, and behavioural consequences that it would be beyond the scope of even an entire thesis to go over the myriad ways that social groups influence behaviour. It is impossible to do full justice to the vast scope of research that has been conducted over

the last 70 years, and so instead - inspired by Van Vugt and Park (2009) - I will highlight just a few select examples of 'key findings' in the study of intergroup behaviour to set the stage for my later discussion. This of course comes with the caveat that this list is necessarily non-exhaustive and intentionally biased towards those findings that bear, to a greater or lesser degree, on intergroup prosocial behaviour.

First, humans have a remarkable tendency to make spontaneous categorizations between the ingroup ("us") and the outgroup ("them") - even when these group memberships are based on arbitrary (e.g. preferences for modern art) or explicitly random criteria (e.g. a coin toss) - and individuals will show preferential treatment for members of their own group on the basis of such categorizations (Brewer, 1979; Tajfel & Turner, 1979; Turner et al., 1987).

Second, humans appear to be unique in their capacity to form deep emotional attachments to groups that are, in many ways, largely symbolic (e.g. sports team, college major). When people highly identify with their group they feel connected to other ingroup members (Doosje, Ellemers, & Spears, 1995), remain committed to their ingroup when faced with threat (Ellemers, Spears, & Doosje, 1997), and are concerned about how their group is treated relative to other groups (Tropp & Wright, 1999). When people identify with a sports team, for example, they experience pleasure when their team does well and this can be seen at the neural level (e.g. Cikara, Botvinick, & Fiske, 2011).

Third, intergroup conflict is characterized by a sense of moral superiority: "we" are more honest, peaceful, trustworthy, and friendly than "them". As the ingroup becomes larger, older, and more established, the norms, rules, and institutions that

maintain cooperation often take on the character of moral authority, and this can then lead to denigration and contempt towards outgroup members (Brewer, 1999; Castano, 2008; Leidner, Castano, Zaiser, & Giner-Sorolla, 2010).

Fourth, humans place a strong premium on loyalty within the group. We expect ingroup members to be more cooperative towards us than outgroup members (Brewer, 2008), and people exhibit a *black sheep effect* whereby deviant ingroup members are judged more negatively than outgroup counterparts (Marques, Yzerbyt, & Leyens, 1988). And in economic games, noncooperative ingroup members are published more harshly than noncooperative outgroup members (Goette, Huffman, & Meier, 2006; Mendoza, Lane, & Amodio, 2014; Shinada, Yamagishi, & Ohmura, 2004) [2].

Fifth, intergroup contexts are often automatically perceived as competitive and hostile. Simply representing a social interaction as 'intergroup' undermines trust in others (Insko & Schopler, 1998; Pemberton, Insko, & Schopler, 1996; Wildschut, Pinter, Vevea, Insko, & Schopler, 2003), and the added perception of competitive relations between groups increases conflict (Campbell, 1965; Sherif, 1966).

Prosocial Behaviour

The kind of intergroup bias I am concerned with in this thesis is that observed in the context of prosocial behaviour. Let me begin by defining my terms. A prototypical

[2] Though it should be noted that other work suggests that in-group members may be punished *more* for selfish behaviour than outgroup members, indicating parochialism in so-called 'altruistic punishment' (Bernhard, Fischbacher, & Fehr, 2006; Mussweiler & Ockenfels, 2013; Yudkin, Rothmund, Twardawski, Thalla, & Van Bavel, 2016). This asymmetry is interesting but not critical to my thesis, which is focused on ingroup-favouring prosocial behaviour rather than punishment per se.

definition in social psychology is that prosocial behaviour refers to "a broad category of acts that are defined by some significant segment of society and/or one's social group as generally beneficial to other people" (Penner, Dovidio, Piliavin, & Schroeder, 2005, p. 366). Breaking this down further, it is argued that prosocial behaviour can be *unidirectional*, where one person provides assistance to another (e.g. helping in emergency situations, volunteering, and donating to charity) and *bidirectional*, where individuals are aware of the benefits of pursuing the best joint outcomes for all and coordinate their behaviour accordingly (e.g. cooperation in interdependent economic games) (Dovidio, Piliavin, Schroeder, & Penner, 2006). As suggested by these definitions, exactly what constitutes prosocial behaviour is quite wide-ranging and context-dependent. Luckily – and as I shall discuss next - in my thesis I am specifically concerned with studying prosocial behaviour using economic games, and these offer a precise operationalization of prosociality: giving up some of ones' own money to someone else. It is therefore this definition that I adopt in this thesis.

For the work reported in my thesis I have decided to use giving in economic games as my primary method, and it may be helpful to give some background for what these are and why I have chosen to use them. Experimental economic games have a long tradition in social psychology, being influenced by work conducted in game theory (Von Neumann & Morgenstern, 1944), but also social psychological work on characterizing social encounters in terms of their potential to promote cooperation or competition (Deutsch, 1949), and understanding behaviour as the function of the personal and the situation (Lewin, 1936). From this, social psychologists have considered *interdependence*

theory (Kelley et al., 2003; Kelley & Thibaut, 1978; Thibaut & Kelley, 1959) and the way that social situations with varying levels of mutual dependence shape both intrapersonal and interpersonal processes.

Why use Economic Games?

Economic games are an effective tool for studying prosocial behaviour for at least five reasons (Pruitt & Kimmel, 1977). First and foremost, they allow researchers to look at behaviour - whether people *actually* make prosocial decisions - and not just whether people *say* that they would do. Second, they permit precise measurements of the extent of cooperation (operationalized by amount of money or percent of endowment) that allow comparisons across studies. Third, they are usually fairly easy and economical experiments to run[3]. Fourth, many sources of variance found in naturalistic settings are more likely to be absent in economic games, reducing the impact of confounding variables. And fifth, competitive or hostile behaviour can be manifested in these games without lasting injury to people or their relationships (Pruitt & Kimmel, 1977).

The first and second advantages that Pruitt and Kimmel identify are particularly important. It is it has long been known that there can exist an important gap between self-reported and actual behaviour (Mischel, 1972). For example, there is often a stark difference between hypothetical and real moral decisions (FeldmanHall et al., 2012) and critically for studying prosocial behaviour, how much people say they would help and how much they actually help are often different things. In a classic study looking at the

[3] Of course, this is a more of a practical than theoretical scientific advantage and should certainly not be a primary reason to use economic games.

effects of physical attractiveness on helping, West and Brown (1975) looked at prosocial behaviour extended to an experimenter who was asking people on the road for money to help pay for medical care at a nearby clinic. Of relevance to this discussion, the authors looked at helping behaviour in two ways: first by asking people what they would do in the situation; and second by actually staging the event. Perhaps unsurprisingly, most participants said that they would give generously. In actuality, they helped much less. One of the benefits of using economic games where participants make real decisions about real allocations of money, then, is that it allows researchers to look at how much participants actually *do* help, and not just how much they think they would help.

Interdependence Theory

Interdependence theory (Kelley et al., 2003; Kelley & Thibaut, 1978; Thibaut & Kelley, 1959) is strongly connected with the use of economic games and has roots both in Lewinian ideas of the *person* x *situation interaction* (Lewin, 1936) and classic game theory (Von Neumann & Morgenstern, 1944). Interdependence theory is a framework for understanding behaviour in interdependent contexts like economic games, and assumes that interactions are the function of an interdependent structure of the situation (e.g. the prisoner's dilemma), the interacting partners (e.g. partner A and B), and interaction dynamics (e.g. the use of a tit-for-tat strategy) (Van Lange, Joireman, Parks, & Van Dijk, 2013). This is outlined in the SABI model (structure, partners A and B, interaction), which can be represented as $I = f(S, A, B)$, where interaction I is a function of two people A and B and their needs, thoughts, and motives in relation to one another, and the interdependent situation S in which the interaction occurs (Holmes, 2002; Kelley et al.,

2003). Interdependence theory distinguishes between the *given matrix* – objective features of the situation and preferences based on self-interest, and the *effective matrix* – subjective preferences "based on broader considerations including partner's interests, long-term goals, or strategic considerations" (Rusbult & Van Lange, 2008, p. 2055). Individual differences in preferences regarding resource allocation are typically understood as representing *social value orientations* (SVO: Messick & McClintock, 1968). In interdependence theory, *transformation* describes the psychological shift in motivation from given to effective preferences.

To bring this together, four basic assumptions of interdependence theory can be discerned (Van Lange, 2012; Van Lange & Rusbult, 2011). First, the *principle of structure* posits that to understand the psychological processes one must consider the interdependence features of the situation, which can be arranged in a taxonomy based on the a) *level of dependence*, or the degree to which one person's outcomes relies upon their partner; b) *mutuality of dependence*, or whether two partners are equally or asymmetrically dependent on each other; c) *covariation of interest*, or whether partners have cooperative or competitive interests; d) *basis of dependence*, or how partners influence each other through norms, punishment, communication, and so on; e) *temporal structure*, or whether the situation persists over time; and f) *information availability*, or the degree to which partners are certain or uncertain about potential outcomes and their partner's motives. Second, the *principle of transformation* posits that interactions are subject to transformations - a psychological process guided by interaction goals – in which individuals consider consequences of their own and others' behaviour in terms of

outcomes for the self and others, and in terms of immediate and future consequences. Third, the *principle of interaction* refers to the SABI – interaction as a function of two persons (A and B) and the objective features of the situation. The situation might activate particular motives and experiences in an individual, and this dynamic interaction yields particular patterns of behaviour. And fourth, the *principle of adaptation* refers to the assumption that repeated social interaction experiences can yield adaptations that are reflected in relatively stable orientations (e.g. SVO) to adopt specific transformations.

In my thesis – and for a variety of reasons - I will not adopt an explicit framework based on the specifics of interdependence theory[4] and SVO[5]. That is, while I will use games that involve interdependence and draw on the insights of interdependence theory

[4] Interdependence theory is a framework for examining interdependent social behaviour, and in my thesis I have focused also on non-interdependent games. Indeed, one of my most important findings is that ingroup favouritism is observed even in non-interdependent contexts. Moreover, interdependence theory specializes in suggesting a broad and integrative framework with which to begin approaching a given psychological phenomenon. This breadth, however, does come with the corresponding feature that it can be difficult to discern specific, focused, falsifiable hypotheses, and there seems relatively little work focusing on specifically applying together all the key integrative principles of interdependence theory to study intergroup prosocial behaviour. For this reason I draw on ideas inspired by interdependence theory, but do not claim, or even seek, to study intergroup prosocial behaviour through this lens.

[5] First, the focus of my thesis is on when people show ingroup favouring prosocial behaviour, rather than the underlying stable patterns of individual differences in prosocial behaviour generally. Second, an important problem with using SVO for my purposes is that I have also focused on non-interdependent interactions, and there are methodological concerns with using SVO to predict behaviour in non-interdependent contexts. The dominant interpretation of SVO appears to be that decomposed SVO games measure some internal social preference for certain distributions of outcomes – for example, "individual differences in social value orientation are predictive of various donations" (Van Lange, Bekkers, Schuyt, & Vugt, 2007, p. 375). In such cases, the SVO is used as an individual difference measure that is then correlated with some kind of resource allocation (e.g. Van Lange et al., 2007) potentially also with some experimental manipulation (e.g. De Dreu & McCusker, 1997). This appears to be problematic, however, when applied to the DG. In the DG, one makes a unilateral decision about how to distribute some resource, which is functionally identical to SVO decomposed games except that the DG typically has a continuous response (e.g. 0-100%) and SVO games have multiple options (e.g. A, B, or C). To the extent that both the DG and SVO are intended to measure social motives (interdependence theory) or social preferences (behavioural economics), and both involve unilateral resource allocation, it becomes circular to use SVO to explain behaviour in DG-type games – which, for reasons outlined later, I use in two of my three empirical chapters.

where useful, my thesis is not wedded to an interdependence theory framework and is not intended to inform interdependence theory per se. I will, however, adopt a few key insights from this work, and this is why I have very briefly outlined the theory here. The first insight is that behaviour in interdependent contexts involves different processes to behaviour in non-interdependent contexts. This will be important in my thesis when, for example, I compare behaviour in the interdependent and non-interdependent games. The second key insight is that individuals often exhibit consistent preferences for prosocial resource allocation, and these prosocial preferences can be inculcated by repeated patterns of interaction. This will be important when discussing models of intuitive prosocial behaviour.

Types of Economic Games

When it comes to using economic games to study prosocial behaviour, research has typically used a small number of specific games with slight modifications. Canonical economic games include the *Dictator Game*[6] (Forsythe, Horowitz, Savin, & Sefton, 1994; Kahneman, Knetsch, & Thaler, 1986); the *Ultimatum Game* (Güth, Schmittberger, & Schwarze, 1982); the *Trust Game* (Berg, Dickhaut, & McCabe, 1995); and social dilemmas such as the *Prisoner's Dilemma* (PD: Axelrod, 1980; Rapoport & Chammah, 1965)*; Common Pool Dilemmas* (CPD: G. Hardin, 1968; Messick & Brewer, 1983); and *Public Goods Dilemmas*

[6] Despite being called the Dictator *Game*, the DG it is not formally a game as defined in economics because to be a game a player's outcome should depend on the actions of another. In this way, the dictator game can be seen as a "degenerate" game. Nonetheless, in the majority of reviews on "economic games" the DG is typically described, interpreted, and used as an economic game and I shall adopt this convention here (Balliet, Wu, & De Dreu, 2014; Camerer, 2003; Camerer & Fehr, 2004; Fehr & Schmidt, 2006)

(PGD: G. Hardin, 1968). In this section I shall very briefly outline the three canonical games that I use in this thesis.

The Prisoner's Dilemma

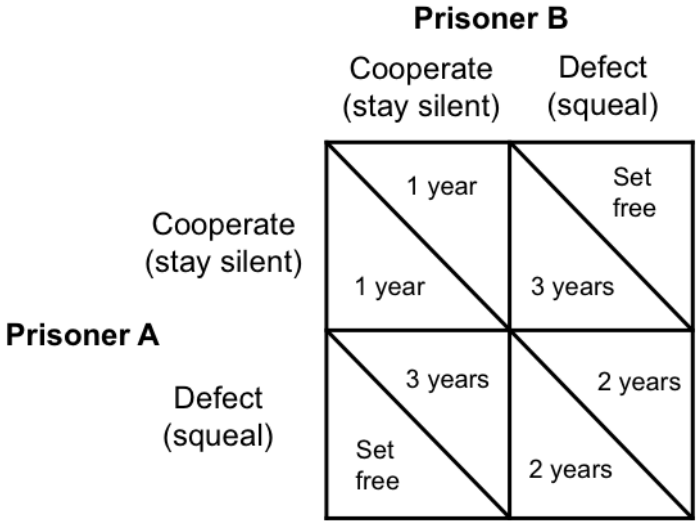
Social dilemmas – situations in which collective interests are at odds with private interests – constitute one of the most common ways to explore cooperation because they exemplify the tension that exists between individual rationality and group rationality (Dawes, 1980; Hardin, 1968; 1998). While the specific characteristics of social dilemmas – both in real life and in research – vary across games and contexts, social dilemmas have two fundamental characteristics: first, that each individual receives a higher payoff for defecting from what is in the collective interest (e.g., using all the available resources for your own advantage) than for cooperating, regardless what other individuals do; and second, that overall everyone is better off if they all cooperate than if they all defect (Dawes, 1980)[7]. In this way, social dilemmas are mixed-motive situations in which there is a conflict between what the player would like to happen and what others would like to happen (Komorita & Parks, 1995).

The roots of experimental social dilemmas lie in the 2-person *Prisoners Dilemma* (PD). The original Prisoner's Dilemma involves a hypothetical situation where two prisoners are in custody for committing a crime. The prosecutors lack sufficient evidence to convict the pair on the principal charge, but could convict them on a lesser offence. The

[7] It should be noted that in a recent integrative review, Van Lange, Joireman, and Parks (2013) argue that while Dawes' definition addresses the crux of the dilemma, it does not fully do justice to other interdependence structures (e.g. the chicken game) and the temporal dimension (short term vs. long term consequences). Van Lange et al. therefore adopt a more inclusive definition of social dilemmas, but this issue is not critical to my experimental work here, which focuses only on a classic PD.

prosecutors offer each prisoner a bargain: they can defect and betray each other by testifying the other committed the crime, or they cooperate and stay silent. As can be seen in **Figure 1.1**, if players both cooperate (stay silent), they achieve a fairly good outcome and are convicted only on the lesser charge (1-year sentence). However, if one player defects (squeals) while the other cooperates (stays silent), the defector receives the best payoff (they are let free), and the cooperator gets the worst payoff (they get a 3-year sentence). Consequently, while both players have an incentive to defect (squeal), the catch is that if both defect, both do poorly (both get a 2-year sentence) (Rapoport & Chammah, 1965).

Figure 1.1: The Prisoner's Dilemma



The Dictator Game

In the *Dictator Game* (DG), one player - the dictator - makes a unilateral decision about how to divide an amount of money with a second player - the recipient (Forsythe et al., 1994; Kahneman et al., 1986). The dictator can allocate any amount of money to the

recipient - from nothing to the entire amount - and the recipient must accept this amount. The DG is therefore a non-interdependent game because the outcomes of the dictator do not depend on the actions of the recipient (because the decision is unilateral). While a self-interested dictator in the DG should transfer nothing, this rarely happens (Camerer, 2003), with a recent meta-analysis of 41,433 DGs showing that around 64% of participants give at least something, and participants give on average 28% of the endowment (Engel, 2011).

The classic interpretation of behaviour in the DG is that it reflects a social motivation to improve the welfare of others (what behavioural economics call *social preferences*). In one prototypical statement: "Dictator games measure pure altruism" (Fehr & Schmidt, 2006, p. 623). It should be noted, however, that within behavioural economics this conclusion is far from uncontroversial. Criticisms have been raised to the effect that behaviour in the DG is an experimenter demand effect driven by perceived identifiability of decisions (Hoffman, McCabe, Shachat, & Smith, 1994)[8] and/or context-dependent

[8] The identifiability critique centres around the idea that giving is merely an experimenter effect resulting from the experimenter knowing the individual's decision decisions (Hoffman, McCabe, Shachat, & Smith, 1994). Hoffman and colleagues (1994) conducted double-blind anonymous DGs and found that 46% donated at least $3 (out of $10) when there was anonymity between participants, but when anonymity between the experimenter and the participant was added, only 16% donated at least $3. Based on such results, they argue that "behavior [in the DG] may be due not to a taste for 'fairness'…but rather to a social concern for what others may think, and for being held in high regard by others" (p.372). On the other hand, that other researchers have found small effects of anonymity. For example, one study showed that introducing double-blind anonymity decreased giving in the DG from 19% to 16% of endowment, but this difference was not statistically significant (Barmettler, Fehr, & Zehnder, 2012). In a recent meta-analysis, Engel (2011) found no evidence that double-blind procedures have a significant effect on behaviour in the DG, at least in one-shot interactions (though there was a small effect of double-blind procedures in iterated games).

social norms[9] (Bardsley, 2008; Levitt & List, 2007; List, 2007). That said, this is not critical to my thesis because it is primarily a debate that occurs within the field of behavioural economics, and is focused primarily around the existence of social preferences. For my purposes, even if behaviour in the DG is affected by identifiability and social norms, it is still a task that unambiguously, efficiently, and precisely measures prosocial behaviour as defined in this thesis: giving up some of one's own money to someone else.

The Trust Game

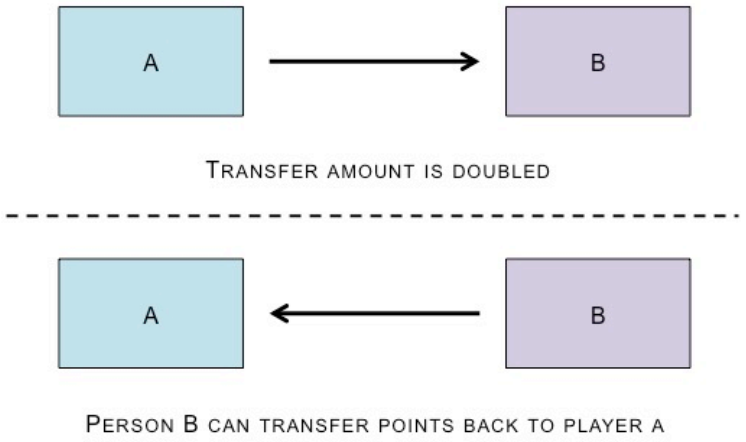
A classic *Trust Game* (Berg, Dickhaut, & McCabe, 1995) has two participants: an investor and a trustee. The investor is given some money and told that they can send a proportion (from zero to the full amount) of this money to the trustee, and that the experimenter will multiply the money by some amount. Once the trustee receives the money, they are told that they can send back a portion of it to the investor, again ranging from zero to the full amount[10] (See **Figure 1.2**).

---

[9] The social norms critique posits that giving in the DG is an experimental artefact to do with social norms and perceptions of the expected behaviour. Bardsley (2008) and List (2007) both conduct similar experiments where, in addition to the giving option where participants choose how much of the endowment to give to the recipient, there is also a taking option where participants can choose how much to take from the recipient. When a taking option is included, fewer participants give money (but participants are still not entirely selfish). In the standard DG, the actions allowed (give vs. keep) might implicitly 'demand' the DG to give some money since not doing so would be to act entirely selfishly with the money that the experimenter has kindly endowed. In contrast, when a taking option in the DG is introduced this changes the social norms such that now the (more) selfish option is to take, and so giving nothing is perceived as morally acceptable.

[10] The TG therefore shares important similarities with both the DG and the PD. The TG can be considered "essentially a dictator game in which the Trustee dictates an allocation, with the difference, however, that the Investor's initial investment determines the amount to be shared" (Fehr & Schmidt, 2006, p. 624). While some have suggested the TG to be a sequential 2-person PD with multiple options (List, 2007), in the PD the risk is symmetric while in the trust game the risk is assumed by the investor. Therefore, the PD could be considered a two-sided trust game where each player is both an investor and trustee (Evans & Krueger, 2009).

Figure 1.2: The Trust Game



As with other economic games, people consistently act in a prosocial way and transfer money. In the original study, 30 out of 32 trials resulted in the investor sending money, with an average of just over 50% of the original endowment, and the average amount returned to the first player was in excess of the amount originally sent (Berg et al., 1995). This is supported by a recent meta-analysis looking at 162 replications of the TG with more than 23,000 participants: the investor transferred an average of 51% of the endowment, and the trustee returned an average of 37% of the money sent (Johnson & Mislin, 2011).

Interpreting what behaviour in the TG means has not been as contentious as for the DG, but some points bear noting. First, despite initial assumptions (Berg et al., 1995), behaviour of the first mover (investor) does not solely measure expectations of reciprocity – though such expectations will naturally be stronger for the investor than the trustee.

For both players in the game, behaviour in the standard TG design can be driven both by trust and expectations of reciprocity, and by relatively more prosocial other-regarding motives (Cox, 2004).

Gains and Losses in Interpersonal Decision Making

The final piece of background information relevant to prosocial behaviour and the empirical work reported in my thesis is the finding that whether a choice is framed in terms of gains or losses has important effects on decision making. Prospect theory (Kahneman & Tversky, 1979) is an influential theory in behavioural economics that looks at how people evaluate risks in decision settings. Standing against the then-dominant expected utility theory, Kahneman and Tversky demonstrate that people often act in ways inconsistent with expected utility, and introduce prospect theory as a model of decision-making under risk that better accounts for behavioural data. One such phenomenon that has been studied under the aegis of prospect theory is loss aversion: the finding that a loss has a stronger influence on choices than a gain of the same magnitude (Kahneman & Tversky, 1979). According to prospect theory, choices are evaluated against a reference outcome (e.g. one's current assets) and then choice outcomes that are above this reference point are evaluated as gains, and choice outcomes

above this reference point as losses. For this reason, framing a choice in terms of gains or losses can have important effects on decisions[11].

The most famous demonstration of loss aversion in risky choice[12] comes from what is known as the "Asian disease problem" (Tversky & Kahneman, 1981). Here, participants are told to:

> "Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. One possible program to combat the disease has been proposed. Assume that the exact scientific estimate of the consequences of this program is as follows…"

After this, some participants are presented with two options (A and B):

A: If this program is adopted, 200 people will be saved.

B: If this program is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

---

[11] There are other, similar, types of framing effects but the one I am interested in is gains and loss framing. Levin, Schneider, and Gaeth (1998) identify three types of framing effects: whether something is framed as positive vs. negative (e.g. 80% lean vs. 20% fat); whether the goal is framed in terms of the benefits of achieving it versus it the problems of not achieving it (e.g. benefits of exercise vs. health problems of obesity); and whether something is framed in terms of gains vs. losses. It is the latter that I am interested in here.

[12] One can, on prospect theory lines, posit risk as a necessary concept with which to understand framing effects in social dilemmas. However, in line with other scholars looking at framing in prosocial decisions I shall focus just on framing and loss aversion rather than risk and uncertainty per se (McCusker & Carnevale, 1995). Loss aversion occurs even in riskless choice (Tversky & Kahneman, 1991) and one can explain framing effects in loss aversion without invoking risk and uncertainty - themselves huge topics of research that are not necessary to delve into here.

Other participants are presented with a different two options (C and D):

C: If this program is adopted, 400 people will die.

D: If this program is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Options A and C are structurally identical, as are Options B and D. For example, if there are 600 people and 200 will live (Option A) this of course means that 400 will die (Option B). Similarly, if there is a 1/3 possibility that everyone will be saved (Option B), this is structurally identical to a 2/3 possibility that everyone will die (Option D). And yet, people do respond drastically differently when presented with these options. Presented with a choice between A and B, around 72% of people choose Option A. Presented with a choice between C and D, however, around 78% of people choose option D. The way that the problem is framed in terms of gains (lives saved) versus losses (lives lost) has a drastic influence on behaviour, with people being more sensitive to potential losses than potential gains.

Loss aversion is a robust phenomenon, being observed in a variety of diverse contexts including tax evasion (O. H. Chang, Nichols, & Schultz, 1987) and clinical health care (McNeil, Pauker, Sox, & Tversky, 1982), and is supported by meta-analytic work looking at 136 papers with almost 30,000 participants and 230 effect sizes (Kühberger, 1998). While prospect theory provides a formal mathematical account of how framing leads to shifts in value compared to a reference point, it arguably (at least originally) has fairly little to say about the cognitive processes that underlie such gains/loss framing

effects and loss aversion (Whitney, Rinehart, & Hinson, 2008). In more recent years, framing effects have been explained in terms of a dual process model[13] whereby the intuitive, heuristic-based response is driven by a negative affective reaction to sure losses but this can be overcome by deliberative, reasoned, processes (Kahneman & Frederick, 2007). Indeed, this is supported by neuroscientific data suggesting that framing effects are associated with activation in areas like the amygdala (Gonzalez, Dana, Koshino, & Just, 2005), which play an important role in emotion processing. Therefore, like the more general 'bad-is-stronger-than-good' bias in human cognition (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), loss aversion might occur because people expect the negative emotional consequences (i.e. pain) of losing something to be greater than the positive emotional consequences (i.e. happiness) of gaining something (e.g. Kahneman & Tversky, 1984).

Why, however, would people be more averse to losses than gains? And why does it feel worse to lose than to gain? One way to understand this is to consider the evolutionary context in which our ancestors lived. In our Pleistocene history in which humans often lived 'on the edge', losing resources (e.g. food) could lead to starvation and even death, while gaining equivalent resources would have only a more short-term relief and respite from the looming spectre of starvation. Survival is more important than prosperity, and given this, it would not be surprising if evolutionary pressures facilitated greater sensitivity to losses than to gains (Chen, Lakshminarayanan, & Santos, 2006;

[13] For a longer discussion about dual process models see page 105 of this thesis.

Hintze, Olson, Adami, & Hertwig, 2015; Li, Kenrick, Griskevicius, & Neuberg, 2012; McDermott, Fowler, & Smirnov, 2008). Such an evolutionary-based account can help to explain why loss aversion is observed in many different cultures (e.g. Bontempo, Bottom, & Weber, 1997)) and even in non-human primates like capuchin monkeys (e.g. Chen et al., 2006). Moreover, recent computational modelling has suggested that risk averse strategies like loss aversion are beneficial as adaptations to living in small groups, and consequently emerge only in groups comparable to the size of our ancestral tribal groups (Hintze et al., 2015).

Inspired by evolutionary theorizing about decision biases, it has been posited that loss aversion should differ depending on the motivations and contexts in which the choice occurs (Li et al., 2012). One such context in which loss aversion be manifested differently is whether choices affect the self or another person. Prospect theory (and the concomitant work on loss aversion) was originally designed for individual decision making, and much work since its formulation has focused on decisions for the self. For the purposes of my thesis, however, it is an interesting question whether differential responses to gains and losses manifests in the context of prosocial behaviour, and particularly intergroup prosocial behaviour. That is: how do participants make costly decisions to help others gain, or avoid losing, resources? There are two lines of reasoning when considering loss aversion in interpersonal contexts: one taking a self-oriented perspective such that people are primarily concerned with minimizing their own losses; and one taking an other-oriented perspective such that people find it more wrong to inflict losses on others than to withhold equivalent gains.

A self-oriented perspective on loss aversion would suggest that, because people are more sensitive to losses than gains, even when interacting with others and even at the expense of the other person, people will be more motivated to minimize their own losses than to maximise gains. Indeed, in negotiation contexts, people demand more in loss frames then gains frames (De Dreu, Emans, Vliert, & Carnevale, 1994; e.g. De Dreu, Emans, & Van de Vliert, 1992). Of most relevance for a self-oriented perspective, Brewer and Kramer (1986) looked at the decision framing by comparing two versions of an n-person PD: a "give-some" loss-framed Public Goods Dilemma (PGD) and a "take-some" gains-framed Common Pool Dilemma (CPD). While give-some dilemmas like the PGD are framed in terms of losses (one has an initial pool of money and then one loses some of the initial pool by giving to the common pool), take-some dilemmas like the CPD are framed in terms of gains (participants start with nothing and there is an existing common pool from which the individual chooses how much to take). Kramer and Brewer utilized this difference, having participants play either a CPD or PGD and experimentally manipulating whether social identity was salient (i.e. people were playing as group members) or not (i.e. people were playing as individuals). They found that people were more cooperative when a social identity was salient, and - consistent with self-oriented loss aversion - were more prosocial in the gains-framed CPD than the loss-framed PGD. Brewer and Kramer (1986) interpret this as suggesting that people in the PGD are more averse to the risk of losing the money they already have, and this individual-level loss aversion then reduces prosocial behaviour. Similar results are found, for example, by McCusker and Carnevale (1995). This is not the only perspective one can take, however.

An other-oriented perspective on loss aversion focuses less on how gains and losses frames influence perceptions of outcomes for the self, and rather on how gains or losses *for the other person* influence prosocial behaviour. Baron (1995, 1996) has argued that when it comes to making decisions involving others, people adopt a heuristic-based *do-no-harm principle*: do not take actions that lead to a worse outcome for someone than that which would have occurred with no action. For example, people are reluctant to endorse vaccination decisions when, although more lives would be saved from the disease as a result, there is a small chance that some children will die from the vaccine (Ritov & Baron, 1990). And there is evidence that loss aversion can encourage prosocial behaviour. For example, van Beest and colleagues (2005) looked at the do-no-harm principle and coalition formation in gains and losses frames and found evidence that people were more motivated by fairness, and less motivated to maximise their outcome, for losses than gains. Similarly, Leliveld, van Beest, van Dijk, and Tenbrunsel (2009) looked at how gains and losses affected motivated bargaining behaviour[14] in the UG and found across three studies that fairness was more cognitively accessible when bargaining occurred in a losses than gains frame, and that a norm of fairness was stronger in losses frames than gain frames. Overall, it seems that because inflicting a loss is seen as more harmful and fairness-violating than withholding a gain, individuals are more likely to

[14] A motivated bargaining approach suggests that to understand negotiation, one should understand the underlying motives of the negotiator. In particular, a motivated bargaining approach recognizes that there are motives both for self-interest and a concern for others (e.g. De Dreu, Beersma, Stroebe, & Euwema, 2006).

help another avoid experiencing a harmful outcome than they are to help provide a positive outcome (Baron, 1995, 1996; Leliveld et al., 2009; Van Beest et al., 2005).

In the context of social dilemmas specifically, De Dreu and McCusker (1997) have sought to integrate the self-oriented and other-oriented findings by considering how prior social motives to cooperate might interact with decision frame in interdependent contexts. Drawing from interdependence theory and work on SVO the idea that for some people ("prosocials") acting prosocially yields a higher value than for others ("proselfs"), and from loss aversion the idea that losses loom larger than gains, they combine these perspectives to suggest that framing outcomes in terms of losses increases the difference in value from acting prosocially or selfish, and therefore, the tendency to prefer acting either prosocially or selfishly – depending on prior motives. That is, "framing outcomes as losses rather than as gains should increase the preference for defection in individualists and competitors and the preference for cooperation in individuals with a cooperative motive."(De Dreu & McCusker, 1997, p. 1095). Indeed, in three studies they provide evidence that whether a decision is framed in terms of gains or losses does seem to transform social motives such that prosocials cooperated more under a loss rather than a gain frame and proselfs cooperated less under a loss than a gain frame.

Overall, there is good evidence that whether a decision is framed in terms of gains and losses has a marked influence on prosocial behaviour. Perhaps surprisingly, then,

almost no work has looked at gains and losses in the context of ingroup-favouring prosocial behaviour[15]. It is this issue that I will address in my first two empirical chapters.

Dual Process Models of Prosocial Behaviour

In the last decade, there has been an increasing focus in applying dual process models of cognition to prosocial behaviour in order to explore whether prosociality requires exerting control or instead comes from more automatic processes. Dual processes models conceptualize cognition as resulting from the competition between quick, intuitive, and automatic processes, and slow, deliberative, and controlled processes (Bargh & Chartrand, 1999; Chaiken & Trope, 1999; Shiffrin & Schneider, 1977; Sloman, 1996; Stanovich & West, 2000). Befitting its status as a key topic across many areas of psychology, these different processes have been used in theories across the field and are referred to by different names – for example, implicit vs. explicit (e.g. Sun, Slusarz, & Terry, 2005), system 1 vs. system 2 (e.g. Stanovich & West, 2000), automatic vs. controlled (e.g. Shiffrin & Schneider, 1977), and intuitive vs. deliberative (e.g. Rand, 2016). In this thesis, I will usually refer to the two processes as intuitive vs. deliberative, but regardless of what they are called, the underlying idea remains the same: cognition can be driven both by cognitive processes that are more implicit, automatic, low-effort, evolutionarily-old and cognitively-quick; and by processes that are more explicit, controlled, high-effort, evolutionarily-recent and cognitively-slow.

[15] Brewer and Kramer (1986) look at the effects of social identity salience, but there is no ingroup/outgroup manipulation, so it is impossible to make any conclusions about ingroup favouritism specifically.

Applying dual process models to the study of prosocial behaviour specifically, it been argued that evidence from neuroscience, developmental psychology, and behavioural studies demonstrates that rather than requiring reflective control, "prosocial behaviour appears to stem from processes that are intuitive, reflexive, and even automatic" (Zaki & Mitchell, 2013, p. 466). How can we tell this? When it comes to differentiating behaviours that are driven more by intuitive or deliberative processes, there are at least three markers.

A first marker of intuitive behaviour comes from neuroscience. Research from imaging techniques such as fMRI has demonstrated that while intuitive processes engage brain regions associated more with value and reward processing such as the ventral striatum and the ventromedial prefrontal cortex, deliberative processes engage brain regions associated with cognitive control and planning such as the lateral prefrontal cortex and the anterior cingulate cortex (Lieberman, 2007). Supporting the claim that prosociality specifically is rooted in intuitive processes, neuroscientific work suggests that that the experience of reward underlies prosocial tendencies (see Zaki & Mitchell, 2013). For example, the ventral striatum and ventromedial prefrontal cortex are associated with observing fair outcomes (Tricomi, Rangel, Camerer, & O'Doherty, 2010), seeing others receive rewards (Zaki, López, & Mitchell, 2014), and when making prosocial decisions oneself (Dawes et al., 2012; Zaki & Mitchell, 2011).

A second marker of intuitive behaviour comes from developmental studies, where intuitive behaviours emerge earlier than controlled ones and before children can effectively exercise cognitive cognition (Harnishfeger, 1995). And, again, evidence

supports prosocial behaviour emerging quickly in developmental ontogeny, with children under one year favouring a prosocial over a selfish agent (Hamlin, Wynn, & Bloom, 2007) and children of 18 months old spontaneously exhibiting helping (Warneken & Tomasello, 2006).

The third marker – and the one that I focus on in my empirical work in Chapter 3 - comes from behavioural studies: while deliberative behaviours occur slowly and are disrupted by cognitive load and distraction, intuitive behaviours are faster and less disrupted by cognitive load and distraction (Chaiken & Trope, 1999; Shiffrin & Schneider, 1977). Given this hallmark of intuitive processes being relatively fast compared to their deliberative counterparts (Kahneman, 2011), a widely-used way to distinguish intuitive vs. deliberative processes is through speed - both as a manipulated and correlational variable.

If decisions resulting more from intuitive processes are made faster than those resulting from deliberative processes and prosocial decisions are made quicker than self-interested ones, this is indicative of prosocial decisions being driven more by intuitive, automatic processes. Indeed, in the paper that kick-started a (renewed) interest in whether people are intuitively cooperative or predisposed to selfishness, Rand, Greene, and Nowak (2012) demonstrated a relationship between time taken to make a decision and the incidence of prosocial behaviour such that the faster a decision is made, the more

likely it is to be cooperative[16] (Studies 2-5). This is instructive evidence but does face the problem that, without an experimental manipulation, it is difficult to determine whether people who made decisions quicker are systematically different in some way from those who make slow decisions, thus confounding results.

Recognising this issue, later studies (both in the original paper by Rand et al. (2012) and other subsequent publications) have focused on the experimental manipulation of time constraint. Because forcing people to decide quickly reduces their ability to deliberate and gives them less opportunity to override automatic, intuitive responses, decisions made under time pressure can be viewed as relatively[17] more intuitive than those made under time delay where people are forced to stop and deliberate (e.g. Kahneman, 2011; Rand et al., 2012). Following this logic, Rand et al. (2012) used a between-subjects design and had participants play a PGD where half of the participants experienced time pressure and were asked to make their decision within 10 seconds, and half of participants experienced time delay and were asked to stop and think about their decision for 10 seconds before responding. Rand's results - both online (Study 6) and in the laboratory (Study 7) – demonstrated that people were more cooperative under time pressure than time delay, indicating that "although the cold logic of self-interest is seductive, our first impulse is to cooperate" (Rand et al., 2012, p. 429).

---

[16] In the work on the SSH and the effect of time pressure on prosocial decisions, these prosocial decisions are often referred to as cooperation, typically operationalized as prosocial behaviour in interdependent games (e.g. the PD, TG) and not prosocial behaviour in games lacking interdependence (e.g. the DG). Cooperation is therefore just a specific form of prosocial behaviour, and following from this previous work I shall use the terms interchangeably in this chapter.

[17] Of course, making people respond quickly does not result in purely intuitive responding, and making people respond slowly does not respond in purely deliberative responding – the claim is rather that time pressure vs. time delay lead to intuition or deliberation respectively emerging as the dominant process.

A growing body of research has continued to document such effects. Rand and colleagues (2014) conduct a meta-analysis on time pressure studies conducted in their lab over a period of two years (15 studies, N = 6190) and find that, while there is substantial variation, time pressure increased cooperation relative to time delay. More recently, Rand (2016) performed a meta-analysis using studies that included not only time pressure manipulations, but also those using cognitive load and priming of intuition (67 studies, N = 17,647), and finds that time pressure does increase cooperation across a variety of contexts - though it should be noted, as I return to in my empirical work in Chapter 3, some researchers have failed to replicate this effect of time pressure on cooperation (e.g. Capraro & Cococcioni, 2016; Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014).

More importantly for this thesis, at a theoretical level, the idea that intuition uniformly favours cooperation—even with outgroup members—is at odds with much past evidence illustrating implicit outgroup bias and distrust. Would intuition favour cooperation even towards competitive outgroup members? This is the question I explore in Chapter 3.

Explaining Ingroup Favouritism in Prosocial Behaviour

Having outlined some necessary background on both intergroup relations and prosocial behaviour, I can now arrive at the phenomenon I will focus on in my thesis: that intragroup prosocial behaviour is typically higher than intergroup prosocial behaviour: people tend to act more prosocially towards ingroup members than outgroup members. Such *ingroup favouritism* or *ingroup-favouring prosocial behaviour* is a robust phenomenon. It is observed even in young children (Fehr, Bernhard, & Rockenbach, 2008) and non-

human primates (Waal, Leimgruber, & Greenberg, 2008). Field studies demonstrate ingroup favouritism across a range of real-world cultures and environments (Habyarimana, Humphreys, Posner, & Weinstein, 2007; Whitt & Wilson, 2007) and even artificially created minimal groups exhibit ingroup favouritism in a variety of economic games (Ahmed, 2007; Halevy, Weisel, & Bornstein, 2012; Tajfel, 1970). When our group membership changes, so too do our parochial boundaries of prosocial behaviour (Kramer & Brewer, 1984). Most compelling is evidence from a recent meta-analysis of 212 studies looking at ingroup favouritism in economic games, which has shown that people are more cooperative with ingroup compared to outgroup members ($d = 0.32$) (Balliet et al., 2014). Across multiple games, by many different researchers, using different groups and populations, the behavioural observation of ingroup favouritism has been well documented. Our prosocial tendencies, in other words, are parochial. Why?

I shall focus on two main approaches to answering this question. The first approach focuses on how simply sharing membership of a social group – '*mere categorization*' – can be enough to cause ingroup-favouring prosociality. Here, I shall first discuss behaviour in the Tajfel minimal group paradigm (Tajfel, 1970) and how SIT was developed in order to explain this (Tajfel & Turner, 1979; Turner et al., 1987). After highlighting problems with a SIT account of ingroup favouritism, I will consider evidence that social identity serves a goal-transformation function that transforms self-interest at the individual level to that of the collective (Brewer, 1979; De Cremer & Van Vugt, 1999). The second approach is based more on the strategic value of ingroup favouritism and focuses on interdependence of outcomes, expectations of reciprocity, and helping

ingroup members more than outgroup members as a way of creating and maintaining a positive reputation. Here, I shall focus specifically on the theory of *bounded generalized reciprocity* (BGR: Yamagishi & Kiyonari, 2000), which assumes that people show ingroup favouritism in prosocial behavior because they expect favourable treatment from fellow ingroup members.

Social Identity and 'Mere Categorization'

The first approach to explaining ingroup-favouring prosocial behaviour is based on the idea that social identity processes can be sufficient, even in the absence of any direct functional benefits, to drive ingroup favouritism.

Social Identity Theory

The roots of motivational accounts of ingroup favouritism lie in Tajfel's (1970) *minimal group paradigm*. Tajfel's classic minimal group studies consisted of two stages. In the first stage, participants were randomly and anonymously categorized into two groups (e.g., "Group A" and "Group B"), purportedly because of some trivial criterion such as estimations of dots or preferences in abstract paintings. Specifically, the use of such minimal groups was intended to look at how the simple act of categorization could produce intergroup behaviour, even when there is no "conflict of interest nor previously existing hostility between the groups" and no "rational link between economic self-interest and the strategy of ingroup favouritism" (Tajfel & Turner, 1979, p. 38). After being assigned to these minimal groups, participants took part in what was described as a set of unrelated resource distribution tasks: the Tajfel Matrices.

Figure 1.3: The Tajfel Matrices

Person A (*Your Group*)	**23**	22	21	20	19	18	17	16	15	14	13	12	**11**
Person B (*Other Group*)	**5**	7	9	11	13	15	17	19	21	23	25	27	**29**

MIP / MD **MJP**

Person A (*Your Group*)	**11**	12	13	14	15	16	17	18	19	20	21	22	**23**
Person B (*Other Group*)	**5**	7	9	11	13	15	17	19	21	23	25	27	**29**

MD **MJP / MIP**

The Tajfel Matrices (Tajfel et al., 1970) required individuals to distribute points between other participants who were identifiable only by code number and group membership (e.g., "participant number 34 from Group A"), and participants were informed that after the task was finished, they would receive the total number of points that had been allocated to them by the other participants (and could not allocate points to themselves). In the task, participants received a booklet containing four matrices including the two reproduced in **Figure 1.3**. The numbers in the boxes represented points to be assigned by the participant to others; the top line indicating to points to be given to Person A, and the bottom line indicating points to be given to Person B. Participants could, for example, choose 23 points for the ingroup member and 5 for the outgroup member (upper matrix; left-hand side) or 22 points for the ingroup and 7 for the outgroup (upper matrix, second from left), and so on. No identifying information was given about

either person, other than whether the person was a member of the participant's own group or another group. Each matrix appeared four times with each row of numbers of labelled to indicate whether the participant was choosing between two outgroup members, two ingroup members, or one ingroup and one outgroup member. Choices were then scored to indicate whether participants opted for point allocations that maximised ingroup profit (MIP), maximised joint profit for both groups ('fairness': MJP), or maximised the difference in favour of the ingroup member (MD)[18].

Tajfel's work using these matrices yielded a consistent pattern of results: participants showed intergroup bias by allocating more points to minimal ingroup members than outgroup members (i.e. they maximise ingroup profit) and often did this in a competitive way, seeking to maximise the difference between ingroup and outgroup members even at the cost of overall profit (for a comprehensive and critical review, see Brewer, 1979). These findings are observed even when division of groups is explicitly random (e.g. a coin toss, rather than purportedly on the basis of shared aesthetic preferences), and studies have shown that participants exhibit more ingroup favouritism on the basis of such explicitly random social categorization than when there was aesthetic similarity but no categorization (i.e. where recipients were referred to only by a code number, not a group) (Billig & Tajfel, 1973). In her critical review, Brewer (1979) argues

[18] In the years since Tajfel and colleagues work on the Tajfel Matrices, a number of alternative matrices have been proposed address various problems with the original matrices (see Bourhis, Sachdev, & Gagnon, 1994). I shall not discuss this further here because my point here is not that the Tajfel matrices are a good measure (of course, for my purposes they'd be a terrible measure because they do not measure prosocial behaviour). Rather, I discuss them now merely as the basis for SIT and SCT and will address concerns of interdependence of outcomes in the appropriate place on page 117.

that even when other factors in the minimal group situation appear to influence behaviour, these can be subsumed under the effects of mere salience of the categorization between ingroup and outgroup: "Factors such as interdependence, intergroup similarity, and shared fate all affect the probability that a respondent will be aware of a relevant basis for categorization into groups, which in turn determines the amount of ingroup bias that is evidenced" (p.319). These findings were not easily explicable by the then-dominant functionalist theories based on group conflict (LeVine & Campbell, 1972; Sherif et al., 1961), and so such findings represented, as Tajfel is reported to have said, "data in search of a theory" (Hogg & Abrams, 1988, p. vii).

The answer to Tajfel's "data in search of a theory" was SIT. Specifically, it was argued that participants in this minimal group context were categorizing themselves in terms of the social category provided by the experimenter, and that such a process of categorization (of self and others) accentuated intergroup differences on the only dimension available—the allocation of points. This accentuation of differences, it is argued, is biased in favour of the ingroup not because of some objective conflict over resources but because individuals are deriving their social identity from the social category which embraces the self. Participants need to maintain positive self-esteem, and in this experimental context this can be achieved by favouring the ingroup (and therefore the self) over the outgroup (Hogg & Abrams, 1988). The problem for my purposes, however, is that such allocations do not represent prosocial behaviour, because there is no transfer of resources from the self to another, only distributions of resources amongst others. can SIT explain ingroup favouring prosocial behaviour?

One way that SIT could explain ingroup favouritism is through ingroup love and depersonalization. Here, the idea is that identifying with a group leads to ingroup favouritism because the individual's own interests become more aligned with the interests of the group collective, thus enhancing the desire to help others—just as one would want to help oneself: "cooperation will follow from the shared and mutual perception of ingroup members of their interests as interchangeable" (Turner et al., 1987, p.65). As Simpson (2006) notes, it is this idea that is typically referred to in discussions of intergroup prosocial behaviour that mention SIT: social identity increases cooperation with ingroup members simply by reducing the tendency for actors to draw distinctions between their own payoffs and the payoffs to fellow ingroup members[19]. As Kramer and Brewer (1984) argue in one oft-cited quote:

> "Inclusion within a common social boundary reduces social distance among group members, making it less likely that individuals will make sharp distinctions between their own and others' welfare. As a result, outcomes for other group members, or for the group as a whole, come to be perceived as one's own" (p.1045).

In this way, social identification leads to a transformation of motives (c.f. interdependence theory: Kelley & Thibaut, 1978) As I shall discuss in the next section,

---

[19] E.g. "[SIT and SCT] assume that once people start to identify with their group, their welfare becomes intertwined with the welfare of the group. According to this view, people who identify with their group may engage in activities to help the group even if it would involve making a personal sacrifice" (Van Vugt & Hart, 2004, p. 586); and "The reduced differentiation between one's own and others' outcomes associated with in-group formation [social identity] provides one mechanism for increasing the weight given to collective outcomes in individual decision making" (Brewer, 1979, p. 322); "An increased group identification may reduce the psychological distance between the group members so that they perceive each other as similar in terms of their goals and achievements. Consequently, people will become motivated to achieve positive outcomes for their group rather than for themselves" (De Cremer & Van Vugt, 1999, p. 872)

there is a growing body of evidence that is consistent with this idea. The problem, however, is that this doesn't seem to connect with some of the more fundamental aspects of classic SIT – the metacontrast principle, the focus on making favourable comparisons, and so on.

The second way that SIT can explain ingroup favouritism relies more on the metacontrast principle and is rooted more strongly in the classic formulation of SIT. Here, the critical aspect is that people are not just concerned with helping the ingroup, but improving the standing of the ingroup in comparison to the outgroup. Based on the metacontrast principle, the presence of intergroup comparisons and an outgroup is necessary for ingroup favouritism to occur: it is only in the presence of a salient outgroup that striving for a positive social identity should manifest itself in tendencies to cooperate with ingroup members more than with outgroup members (Balliet et al., 2014; Ellemers & Haslam, 2011). But this seems to stand in contrast with the previous claim that "cooperation will follow from the shared and mutual perception of ingroup members of their interests as interchangeable" (Turner et al., 1987, p.65). It is likely that a social identity theorist could develop an argument that incorporates these seemingly inconsistent arguments (if they haven't already and I cannot find it). But regardless, the available evidence suggests that SIT-proper (i.e. explicitly considering metacontrast and favourable comparisons) does not explain ingroup favouritism as well as other, more recent, models anyway.

According to SIT, preferential treatment of the ingroup – ingroup favouritism[20] – is the mirror image of derogating and even harming the outgroup – *outgroup derogation* (Tajfel & Turner, 1979; Turner et al., 1987). That is, perceptions of similarity within groups are associated with perceptions of dissimilarity between groups, and (at least one source of) positive social identity for one group comes from their relative standing to another group (Tajfel & Turner, 1979). However, a large body of literature now suggests that ingroup favouritism and outgroup derogation are psychologically distinct: they emerge from different structural and psychological processes and have different psychological consequences (e.g. see Brewer, 1999; Hewstone et al., 2002).

A number of studies by Mummendey, Otten, and colleagues have demonstrated a *positive-negative asymmetry effect* (Blanz, Mummendey, & Otten, 1997; Mummendey & Otten, 1998; Mummendey, Otten, Berger, & Kessler, 2000; Otten, Mummendey, & Blanz, 1996) whereby when the resources allocated in the Tajfel matrices are negative (e.g. unpleasant noise or number of tedious tasks) instead of positive (e.g. points or money)

[20] These are sometimes called "ingroup love" and "outgroup hate" (e.g. Brewer, 1999), but this terminology can be confusing. First, it implicitly assumes the psychological motives of love and hate - it is not clear that a given observation of ingroup favouritism is due to "love", or that failing to help outgroup members is "hate". Second, ingroup love in the sense of preferential treatment of ingroup members is identical to ingroup favouritism, but this then causes confusion when referencing accounts that posit ingroup members more than outgroups because they care more – of course, one cannot posit ingroup love (caring about ingroup members more) to explain ingroup love (ingroup favouritism). I therefore adopt the more neutral terms of ingroup favouritism and outgroup derogation.

ingroup favouritism almost disappears[21] (for a meta-analysis, see Buhl, 1999). This is consistent with intergroup bias being driven more by a preference for the ingroup rather than a desire to derogate the outgroup.

Of particular importance is work by Halevy and collagues (2008; 2012), who have explored the role of ingroup-favouritism and outgroup derogation using the *Intergroup Prisoner's Dilemma – Maximizing Difference* game (IPD-MD: Halevy et al., 2008). The IPD-MD is explicitly designed to test between the motives of ingroup favouritism and outgroup derogation. The IPD-MD has two groups, where each group member is given some money and has to decide how much to contribute to one of two pools: one that benefits the ingroup, and one that harms the outgroup as well as benefiiting the ingroup. Using the IPD-MD, Halevy et al. (2008) have shown that people are not competitive or aggressive *per se*, and that participants almost exclusively contribute to the cooperative within-group pool rather than the competitive between-group pool. Similarly, Halevy et al. (2012) show that in repeated interactions, ingroup favouritism prevails over outgroup derogation even after intergroup conflict. Research using alternative PD matrices has shown that ingroup favouritism arises even in the absence of an outgroup and intergroup comparison, suggesting that helping the ingroup is more potent in driving differential

[21] At first blush, this positive-negative asymmetry might appear to track with my earlier distinction between gains and loss frames. The problem is, however, that such work on the positive-negative asymmetry does not measure prosocial behaviour because – like in the original Tajfel matrices - there is no cost to the participant that has a corresponding benefit to the recipient. While this work is excellent at highlighting the difference in positive and negative allocations in intergroup contexts, it does not – and was not intended to – tell us much about loss aversion in intergroup prosocial behaviour. Correspondingly, in the 'key' papers on the positive-negative asymmetry effect, there is no mention of loss aversion or prospect theory, simply because this is not what they were studying (Mummendey et al., 1992; Mummendey, Otten, Berger, & Kessler, 2000; Otten & Mummendey, 1999).

intergroup prosocial behaviour than harming the outgroup (L. Gaertner, Iuzzini, Witt, & Oriña, 2006).

Compelling evidence that SIT is not the best account of ingroup favouritism comes from Balliet, Wu, and De Dreu (2014), who conduct a theoretically motivated meta-analysis of ingroup favouritism in economic games. They looked at 212 studies from 77 articles and tested a series of eight hypotheses that pitted SIT against an alternative model based on the theory of *bounded generalized reciprocity* (BGR). I shall discuss the BGR at length in the next section, but suffice it to say here that a crucial difference between two is that while SIT focuses more on how group memberships change motivational inclincations to help ingroup members and enhance differences between the ingroup and outgroup, the BGR suggests that the critical driver of ingroup favouritism is instead the strategic expectation that ingroup members (more than outgroup members) will act prosocially towards you. In their meta-analysis, Balliet et al. (2014) suggest that while "findings do fit core propositions advanced in SIT, the results fit better a BGR perspective on ingroup favouritism in cooperation" (Balliet et al., 2014, p. 1576). For example, the metacontrast principle suggests that the presence of an outgroup is necessary for ingroup categorization and subsequent discrimination, but their meta-analysis shows not only are the effects of ingroup love are stronger than outgroup derogation, but that the presence of an outgroup is not necessary for ingroup favouritism (Balliet et al., 2014). Similarly, while SIT predicts no difference in ingroup favouritism between simultaneous decisions (e.g. the PD) and sequential ones (e.g. first mover in the TG), their meta-analysis suggests that this does in fact have a strong influence on the extent of ingroup favouritism, with

an average effect size of d = 0.42 in simultaneous dilemmas and d = 0.26 in sequential games. And finally, SIT cannot well explain differences between common and unilateral knowledge (which I will discuss more in the next section). SIT suggests that ingroup favouritism should be observed whether regardless of whether the group identities are unilateral (i.e. I know that you're an ingroup member but you don't know I am an ingroup member) or common knowledge (i.e. I know that the recipient knows that I am an ingroup member) because both provide an opportunity for people to positively discriminate the ingroup from the outgroup. Balliet et al.'s meta-analysis, however, shows that ingroup favouritism is observed under common knowledge (d = 0.32) but not unilateral knowledge (d = -0.04), which is not explicable by the traditional SIT framework.

Overall, there is fairly little evidence that SIT provides a more compelling explanation of intergroup prosocial behaviour than the alternatives. However, one must not throw the baby out with the bathwater. There is a great deal of evidence for the importance of social identity in ingroup favouritism (e.g. Brewer & Kramer, 1986; De Cremer & Van Vugt, 1999; Kramer & Brewer, 1984; Wit & Kerr, 2002) such that sharing a social identity can be sufficient to lead to ingroup-favouring prosocial behaviour - but not necessarily for the reasons that SIT says it does. Even if SIT has serious limitations, an approach that recognises the importance of individual's social identities may provide a necessary piece of the puzzle of why people tend to help ingroup members more than outgroup members even in the absence of 'rational', economic benefits.

Social Identification as Motivation Transformation

One way that shared social identity can lead to ingroup favouritism is through the idea of *transformation of motives* in interdependence theory (Kelley & Thibaut, 1978; Thibaut & Kelley, 1959). Put simply, because increased group identification reduces the psychological distance between group members and encourages a perception of interchangeable goals and interests (c.f. SIT: Tajfel & Turner, 1979; Turner et al., 1987), this can lead to a transformation of motives whereby self-interest at the personal level is redefined at the collective level and the outcomes for self and others in the group become practically interchangeable (Brewer, 1979). In other words, "when the definition of self changes, the meaning of self-interest and self-serving motivations also changes accordingly" (Brewer, 1991, p. 476)

As noted above, it is this idea that is implicit in Turner's statement that "cooperation will follow from the shared and mutual perception of ingroup members of their interests as interchangeable" (Turner et al., 1987, p.65), but differs from SIT proper in that the focus here is on how shared identity transforms one's motives, rather than metacontrasts, positive self-esteem, and favourable ingroup comparisons. In this sense, a transformation of motives argument is based on the concept of social identity rather than social identity *theory* per se.

The evidence reported by Brewer and Kramer, whereby making a superordinate over subordinate group identity salient encourages cooperation in CPD and PGDs (Brewer & Kramer, 1986; Kramer & Brewer, 1984), is consistent with this idea. It does not, however, directly manipulate whether this is because of a transformation in motives or

some other process, like greater expectations of trust from ingroup members. One of the first studies to directly test a motive (goal) transformation hypothesis comes from De Cremer and Van Vugt (1999), who capitalized on the fact that people have pre-existing differences in preferences for outcome distribution between the self and other. People have, that is, different *social value orientations* (SVO: Messick & McClintock, 1968; Van Lange & Liebrand, 1991) such that they can be described as cooperative (concerned with maximising outcomes for the self and others), competitive (concerned with maximising the difference in outcomes for the self and others), or individualistic (concerned with maximising outcomes for the self but with no preference over the outcomes of others) (Messick & McClintock, 1968). In more recent years it has become more common to describe cooperative people as *prosocials*, and to describe competitive and individualistic together as *proselfs* (Van Lange & Liebrand, 1991). People with prosocial orientations are more cooperative than proselfs in experimental social dilemmas (for a meta-analysis, see Balliet, Parks, & Joireman, 2009) and in other contexts like negotiations (De Dreu & Van Lange, 1995) and commuting decisions (Joireman, Van Lange, Kuhlman, Van Vugt, & Shelley, 1997).

Given these differences in SVO, De Cremer and Van Vugt (1999) reasoned that if the effects of social identification are due to a transformation of motives – the *goal-transformation hypothesis* - this should be particularly so for those proselfs who would not normally be willing to contribute. This is because proselfs start with a focus on self-interest, but then – if social identification transforms motives – when group identification is high, the focus of proselfs would shift from personal outcomes to collective outcomes

so that they become more cooperative. In contrast, prosocials should be less affected by social identification because they already care about the welfare of the group and so there is little room for a transformation of motives to occur. This goal-transformation hypothesis is set against a *goal-amplification hypothesis*. The goal-amplification hypothesis is based on the goal-expectation theory of Pruitt and Kimmel (1977), which suggests that trust (i.e. expectations) only increases cooperation if there is a motivation (i.e. goal) to do so: "While usually necessary, the goal of achieving mutual cooperation, is insufficient to elicit cooperative behavior. It must be accompanied by an expectation that the other will cooperate" (p. 375). The idea here is that shared group membership enhances mutual trust and so social identity does not necessarily increase the value of the collective outcome, but rather reduces the fear that others will free-ride. If this goal-amplification hypothesis is true, proselfs should be fairly insensitive to social identification increasing trust because they lack the motivation to cooperate in the first place. Comparison of these two hypotheses, then, sets up opposing predictions: the goal-transformation hypothesis suggests that social identification will increase cooperation amongst proselfs but not prosocials, while the goal-amplification hypothesis suggests that social identification will increase cooperation amongst prosocials but not proselfs.

To test between these hypotheses, the authors conducted three studies in which they measured SVO through the standard decomposed games and manipulated social identity by telling them the focus was either on individual behaviour (low social identity salience) or comparing the group performance to that from another university (high social identity salience). After this, participants played a one-shot PGD. Their results

supported predictions of the goal-transformation hypothesis: when group identification was low, prosocials were more cooperative than proselfs. However, when group identification was high, proselfs contributed at the same level as prosocials. Such findings lead De Cremer and Van Vugt to conclude that in some cases at least, "group identification gives rise to transformation of motivation whereby the distinction between personal and collective interests is blurred" (De Cremer & Van Vugt, 1999, p. 887).

There are, however, at least two limitations of these studies that should be noted. First and foremost, these studies do not actually investigate the process variables that are assumed to underlie the effect – expectations of others' cooperation (goal-amplification) and individual degree of identification (goal-transformation). Given that the strength of social identification determines the impact of identity on cognition and behaviour (Ashmore, Deaux, & McLaughlin-Volpe, 2004), one would predict goal-transformation to be mediated through the individual strength of identification and not just the group-level manipulation. Second, and relatedly, identification in these studies was manipulated by the introduction of social comparison between the ingroup and outgroup. This might, however, activate social comparison motives (c.f. SIT) such that people cooperate more with the ingroup to 'win' this competition rather than because of a depersonalized sense of self.

This issue was addressed by De Cremer and colleagues (2008). Like in the previous studies, they measured SVO and manipulated high vs. low group identification before having participants play a PGD. This time, however, they manipulated group identification by either telling participants that four of the five team members had the

same major as the participant did (high identification) or that all the other team members had a different major to the participant (low identification). Moreover, they added questions concerning the participants' subjective identification with the group (e.g. "I feel a tie with this group") and how much they expected the others to cooperate. Supporting the goal-transformation hypothesis and again replicating earlier findings, there was a significant interaction between group identification and SVO such that proselfs were more influenced by the group manipulation than prosocials. And critically, mediational analyses showed that it was the self-reported identification ("sense of collective self") and not expectations of cooperation from the team members that mediated this interaction between group and SVO. It is, in other words, the subjective importance that individuals attached to the group that mediated the shift of proselfs into cooperators.

Overall, then, there is a good body of evidence for a social identity-based goal-transformation hypothesis: social identity can lead to a transformation of motives whereby self-interest at the personal level is redefined at the collective level and the outcomes for self and others in the group become practically interchangeable (Brewer, 1979). It is important to note, however, that there is nothing in this explanation that suggests an incompatibility with accounts that focus more on shared group membership increasing trust. It is perfectly possible – indeed, likely – that social identity affects prosocial behaviour both through a transformation of motives and through an increase in expectations of reciprocity. This, of course, is what Brewer (1979) herself originally argued and expanded on in later years. In her overview of the key themes she worked on during her life, she emphasised the idea that group membership leads to depersonalized

trust but suggests that this is not enough because without the motivation to cooperate more with the ingroup, an individual can continue to free-ride the benefits that group membership affords without contributing. Instead, she argues, "goal transformation provides a basis for ingroup cooperation that does not depend directly on expectations that others in the group will reciprocate cooperation" (Brewer, 2011, p. 87). Social identity-based goal transformation is a "basis": it is not, and should not be, the whole story.

Reciprocity and Reputation

The second approach to explaining ingroup-favouring prosocial behaviour focuses less on how sharing a social identity might lead to motivational changes that influence how much we value others' outcomes, and instead takes a more strategic view of ingroup favouritism focused on interdependence of outcomes, reciprocity, and reputation. The claim here is that sharing a social identity comes with expectations that ingroup members (more than outgroup members) will act prosocially towards you, and it this this strategic element that makes it advantageous for you to favour the ingroup.

Outcome Interdependence and Direct Reciprocity

An influential early criticism of the 'mere categorization' love account of ingroup favouritism in the Tajfel matrices was proposed by Rabbie, Schot, and Visser (1989), who argued that ingroup favouritism in point allocation was the result of beliefs about *outcome interdependence* rather than any greater motivation to help ingroup members (Rabbie et al., 1989). While there is – to use the terminology of interdependence theory – no behavioural control in the Tajfel Matrices (i.e. one cannot affect another's behaviour and

the choice of allocations by an individual has no consequence for the allocator), there is implicit interdependence of outcomes in the form of mutual fate control (Thibaut & Kelley, 1959). That is, participants implicitly perceive their own outcomes to be dependent on their choices: "by giving more to their in-group members than to the outgroup members—in the expectation that the other in-group member will reciprocate this implicit cooperative interaction—they will increase their chances of maximizing their own outcomes" (Rabbie et al., 1989, p. 176).

Indeed, work conducted in the years after Tajfel's initial studies has documented that while social categorization can be a sufficient condition for intergroup discrimination, this is by no means a universal response, and can be extinguished by feedback concerning how other members of the ingroup and outgroup respond (Locksley, Ortiz, & Hepburn, 1980). Learning that outgroup members do not discriminate against one's own group and that one's own group does not discriminate against the outgroup reduces discrimination.

Similarly, Rabbie et al. (1989) show that when the participants allocated rewards to both ingroup vs. outgroup members but participants' rewards were determined only by outgroup members, participants showed outgroup favouring allocations. According to Rabbie et al., participants make ingroup-favouring allocations due to perceived outcome interdependence: even if participants cannot allocate money to themselves, they can give to others who may in turn reciprocate their allocations. Rabbie et al. suggest that participants perceive stronger outcome dependence with ingroup members, which in turn leads to higher expectations of reciprocal behaviour by ingroup members and it is

this that leads participants to allocate more points to ingroup members. Rabbie et al. concluded that allocations in the Tajfel matrices were not motivated by a motivation to help ingroup members, but rather reciprocal expectations: the *reciprocity hypothesis* (Rabbie et al., 1989).

Scholars have distinguished between two versions of such a theory: a "strong" *unbounded* version and a "weak" *bounded* version (Stroebe, Lodewijkx, & Spears, 2005). In the unbounded version, participants attempt to maximize their outcomes by allocating more resources to others upon whom they perceive themselves to depend for their own outcomes, anticipating that this favourable treatment will be returned (Rabbie et al., 1989). This unbounded version is "strong" because it posits a direct link between actual outcome dependence and expectations of reciprocity, and makes no claim concerning group membership: this can occur for both ingroup and outgroup members. In contrast, a "weaker" bounded reciprocity hypothesis suggests that the effects of reciprocity are constrained by social categorization such that expectations of reciprocity are not only influenced by interdependence of outcomes but also by group membership (L. Gaertner & Insko, 2000). On this account, people have higher expectations of reciprocity from ingroup members and this leads to ingroup favouritism. Applied to the minimal group studies with which it was formulated, the bounded reciprocity hypothesis suggests that participants who are dependent on their ingroup for their own outcomes have stronger ingroup reciprocal expectations and consequently make more ingroup-favouring allocations in the Tajfel matrices than participants who are not. In contrast, outgroup outcome dependence will not lead to outgroup reciprocal expectations or to outgroup-

favouring allocations (Stroebe et al., 2005). Group membership, in other words, moderates the effects of interdependence and reciprocity in the Tajfel matrices.

In a series of experiments using artificial groups, Yamagishi and colleagues have provided evidence that ingroup favouritism in economic games occurs only when participants believe other ingroup members will reciprocate the favour. Initial evidence for this was provided when Karp et al. conducted a partial replication of Tajfel's minimal group experiments using the Tajfel matrices (Karp, Jin, Yamagishi, & Shinotsuka, 1993). In their study, they had two conditions: a replication of the original minimal group studies where the participant making the allocation decision was also subject to allocation decisions by other members; and a modified condition where allocators were told they would be paid a fixed amount of money and that this would not depend upon others' allocation decisions. Their results revealed that ingroup favouring decisions emerged only when the participant was also a target of allocation decisions by others, and not when participants' payoffs did not depend on allocation decisions by others. Such results speak against an account based purely on social identification, which would predict ingroup favouring decisions whether the participant could receive allocation rewards from others or not. In contrast, such results support the role of a distinct expectation-based process: people only show ingroup favouring decisions when they expect the favour to be reciprocated. In a later study, it was found that participants only acted more cooperatively in a PD when they believed that the recipient had knowledge that the participant was an ingroup member, consistent with the notion that cooperation was driven by the expectations of ingroup reciprocity (Yamagishi & Kiyonari, 2000). This is

further supported by work from L. Gaertner and Insko, who had participants allocate rewards in a minimal group paradigm but varied whether the other allocator was an ingroup or outgroup member, and whether participants would personally receive rewards or not. They found that participants favoured the ingroup over the outgroup, but only when they were dependent on an ingroup member for their own outcomes (L. Gaertner & Insko, 2000). Similarly, Stroebe and colleagues orthogonally manipulated participants' dependence on an ingroup and an outgroup member for outcome rewards in a modified Tajfel Matrices task and found again that ingroup favouring strategies were strongest with (but not exclusive to) outcome dependence on the ingroup (Stroebe et al., 2005).

Work on expectations of reciprocity and interdependence, then, supports a model where individuals respond to the dependence structure and then reciprocate with favouritism towards those on whom they are dependent, with this effect considerably stronger for the ingroup (hence "bounded") due to the perception of the group as a "container of generalized reciprocity" (Yamagishi & Kiyonari, 2000). Indeed, a recent meta-analysis of cooperation in ingroup favouritism by Balliet, Wu, and De Dreu (2014) has provided substantial support for this, with situations involving interdependence of outcomes resulting in stronger ingroup favouritism (e.g. PD, TG: d = 0.42) compared to the weaker (but still significant) effect of games with no interdependence of outcomes (e.g. DGs: d = 0.19). It is clear, however, that while important, interdependence of outcomes cannot completely explain ingroup favouritism, for such behaviour is observed even in non-interdependent games. Why might this be so?

Bounded Generalized Reciprocity

Reputational concerns can lead to ingroup favouritism if group members are strategically concerned with signalling a positive reputation towards other ingroup members. That is, individuals may believe that it is advantageous to selectively act prosocially towards ingroup members when they can build a reputation of being a prosocial person. In line with this, in more recent years Yamagishi et al. have expanded on their earlier work that primarily concerned the potential for direct reciprocity (Jin & Yamagishi, 1997; Yamagishi, Jin, & Kiyonari, 1999), to put forward their *Bounded Generalized Reciprocity* (BGR) Model. They suggest that the presence of a salient ingroup activates a default group heuristic strategy, which explains the higher incidence of prosocial behaviour within groups (Yamagishi & Kiyonari, 2000; Yamagishi & Mifune, 2008, 2009).

The BGR model has three core ideas: first, that humans evolved to have a generalized trust that other ingroup members will cooperate; second, that people are motivated to establish and maintain a cooperative reputation among ingroup members because of the strategic advantages this brings; and third, that people expect to be the beneficiaries of prosocial behaviour from other ingroup members, but not necessarily from the same ingroup members they cooperated with or helped previously (Kiyonari & Yamagishi, 2004). According to the BGR model, people treat ingroup members more favourably than outgroup members not because of any real preference or 'transformed motivation' to do so, but because they expect favourable treatment from ingroup members through both direct and indirect reciprocity (Yamagishi & Kiyonari, 2000). Put

simply, group members believe that others in their group are likely to act prosocially towards them, which influences behaviour. Awareness of these generalized norms of reciprocity within groups, Yamagishi et al. argue, accounts for observations of ingroup favouritism, particularly in social dilemmas where one's own payoff depends on the actions of other group members.

Yamagishi and colleagues have collected a large amount of data consistent with the claims of the BGR model. Yamagishi and Mifune (2008), for example, found that participants only allocated more money to ingroup members in a DG when they believed that others knew their own group membership. Similarly, ingroup favouritism in a DG occurs only when participants are exposed to reputation monitoring cues, suggesting that it is adherence to norms of reciprocity and beliefs about indirect reciprocity rather than ingroup love that drives cooperative behaviour within groups (Mifune, Hashimoto, & Yamagishi, 2010). They argue, therefore, that "in-group favouritism *does not* occur when participants cannot or do not expect favourable treatments from in-group members" (Yamagishi & Kiyonari, 2000, p. 127, my italics).

Consistent with this, meta-analytic work suggests that intergroup discrimination in cooperation emerges when both actor and recipient are aware of group membership (153 samples: $d = 0.32$) but not when recipient does not know the group membership of the actor (15 samples: $d = 0.04$) (Balliet et al., 2014). As I shall discuss more in the next section, however, some findings not included amongst these 15 do find that even if weaker, ingroup favouritism does sometimes appear even in private conditions (Güth, Ploner, & Regner, 2009; Ockenfels & Werner, 2014; Rand et al., 2009).

Can we confidently conclude, based on the BGR, that common knowledge is necessary for ingroup favouritism? I suggest not. The first thing to note is that, of course, while Balliet et al.'s meta-analysis finds no overall effect of ingroup favouritism in private knowledge, there are individual studies included in the meta-analysis that do provide evidence against the BGR, just not on the common (public) vs. unilateral (private) knowledge dimension. Simpson (2006), for example, looked at minimal group identity in a sequential PD and found that for the first mover in a sequential PD, there was no significant difference in rates of cooperation with a minimal ingroup vs. outgroup member. This is consistent with the BGR model, where for the first mover the fear motive is decreased and there is no need to rely on a group-based heuristic: simple direct reciprocity motives can suffice instead. However, second movers, even after the first mover had cooperated (and so there was no fear motive), were still twice as likely to cooperate with a minimal ingroup member than outgroup member. This is inconsistent with the BGR model's suggestion that direct reciprocity will override ingroup-based reciprocity, since when the first mover has already cooperated the expectations of direct reciprocity are even more powerful.

Moreover, and regarding the common (public) vs. unilateral (private) knowledge difference, despite their other work in which they are definitive that ingroup favouritism should not occur under private knowledge (i.e. where the recipient does not know the group membership of the actor), in at least one paper Yamagishi and Mifune (2009)

hypothesise[22] – and demonstrate - that even in private conditions men (but not women) show ingroup favouritism. Rather than treating cases where ingroup favouritism emerges in private conditions as an aberration (like BGR does), it remains interesting to consider when – and why - prosocial behaviour would be observed under anonymity. Further, in the meta-analysis there are only 15 samples with private knowledge compared to 153 with common knowledge. This might not ordinarily be a problem, but it does appear that some studies that do find ingroup favouritism in private conditions are not included in this set for various reasons, and if these were included the results could have been different. For example, the meta-analysis includes four samples from Güth, Ploner, and Regner (2009). In their study, they use a DG and look at the effects of minimal group membership on transfer to ingroup and outgroup members, under conditions of anonymity or not. Their results show that ingroup favouritism is strongest is the common knowledge condition, but that even under private knowledge minimal group members transfer more to ingroup than outgroup members in the DG. In Balliet et al.'s meta-analysis, however, all four of these samples are coded as common knowledge, which does not seem to track with what is reported in Güth et al.'s paper. Similarly, the study of Rand et al. (2009) appears to be coded as common knowledge, but (as clarified in personal communication) in Rand et al. (2009) recipients did not know the group membership of the actor, even though this was not made explicit in the paper. A third – and more

[22] This might suggest that even though the BGR model offers very strong claims - "in-group favouritism *does not* occur when participants cannot or do not expect favourable treatments from in-group members" (Yamagishi & Kiyonari, 2000, p. 127), in practice the BGR is used in a weaker sense such that ingroup favouritism should be stronger in, but not exclusive to, common knowledge conditions.

uncertain - example comes from Chen and Li (2009), which is cited in the reference list and *appears*[23] to show ingroup favouritism in DGs with no common knowledge, but is not included in the meta-analysis.

Moreover, since the meta-analysis has been published, another study has been published that also shows ingroup favouritism in the DG under conditions of anonymity (Ockenfels & Werner, 2014). Using real university-based social groups, Ockenfels and Werner compared behaviour in the DG towards ingroup and outgroup members, again under conditions of common or private knowledge. Like Güth et al. (2009) they find ingroup favouritism to be significantly stronger under common knowledge, but still occurred in the private conditions. Were the meta-analysis to be conducted again with these studies included, it is entirely plausible that meta-analytic findings of ingroup favouritism in private conditions would be observed. Given this, I would suggest that it would be premature to conclude, as per the BGR, that ingroup favouritism does not occur under private conditions, but instead that it should be highlighted as a topic of future research. As Balliet and colleagues write themselves: "new research is needed to unravel whether ingroup favouritism in situations lacking interdependence emerges because of social identity concerns, reputational concerns, or perhaps some combination of these two" (Balliet et al., 2014, p. 1572).

---

[23] Chen and Li report a behemoth paper that is in the style of economics, written for economists, and thus does not include some of the information that psychologists might have included. They conduct 36 independent experimental sessions with 24 different games and five different conditions. There is no explicit mention of whether group membership was common knowledge and the experimental procedure they use for the DG *suggests* decisions were private, but I am not certain. It is plausible that this lack of detail was why it was not included in the meta-analysis originally.

An important limitation of the BGR model is that it has focused on ingroup favouritism in minimal, artificial groups. At the empirical level, for example, 13 of the 15 private knowledge samples included in Balliet et al.'s meta-analysis involved experimentally manipulated groups such as in the minimal group paradigm. Moreover, whether the BGR model is intended to explain behaviour just in minimal groups or in other groups too is unclear. While Yamagishi and Kiyonari (2000) write that the empirical findings in support of the BGR likely only apply to the minimal group situation and that "ingroup favouritism in realistic groups can be attributed to many causes" (p.128), the BGR model is often implicitly presented as an explanation for ingroup favouritism generally. For example, Yamagishi and Mifune (2008) write that "identifying oneself with a social category should not play an important role in our lives unless the social category is a source of tangible outcomes" (p.25), and the discussion supporting the BGR in Balliet et al. (2014) is not limited to only minimal groups, but as an explanation more generally.

Individuals often attach much importance to their social groups in real life, and cooperation is commonly found in groups whose members interact on a regular basis, have emotional ties to one another, share a common frame of reference, and are behaviourally interdependent - conditions absent from minimal group paradigms (Levine & Moreland, 1994; Tajfel & Turner, 1979). While a minor form of this ingroup attachment can be generated in minimal group paradigms, these minimal group memberships may not generate sufficient attachment and emotional involvement to consistently lead to a transformation of motivation in artificial experiments. Indeed, people in self-selected real groups are more cooperative in simultaneous dilemmas than

their counterparts in minimal groups, and this effect is mediated by group identity (Jackson, 2008). It is clear that further research is needed to investigate the effectiveness of the BGR model in predicting prosocial behaviour in real groups, where social identity and transformation of motives are likely to be more important.

A persistent concern with the BGR model as currently formulated is that it seems to ignore the role of other non-expectation-based motives. While work on transformation of motives is compatible with studies supporting the BGR because this motivation is one among many, the reverse is not true because the typical formulation of the BGR is exclusionary: "In-group favoritism in [minimal group experiments] thus has been shown not to be a matter of taste for, or liking of, in-group members per se" (Yamagishi & Kiyonari, 2000, p. 118). This means that the BGR model requires stronger evidence, because it must show there is *no* role for social preferences or prosocial transformation of motives. A more promising and conciliatory approach comes from Brewer (2008), who suggests that ingroup favouritism can occur by two paths:

> "One relies on depersonalized trust—the expectation that others in the group will cooperate because we share group membership—supported by established cooperative ingroup norms and potential sanctions for noncooperation. The second path is based on social identity processes and associated goal transformation and projection of one's own cooperative (group-benefiting) motives to fellow group members. No doubt, successful groups of all types rely to some extent on both paths to generate and sustain intragroup cooperation" (p226).

Overview of Chapters

In my thesis, I present three targeted empirical chapters that shed new insight into ingroup favouritism in prosocial behaviour. In each chapter I look at how group membership influences prosocial behaviour in four carefully selected economic games that help me gain insight both into the specific question of when and why individuals show intergroup bias in prosocial behaviour, and more general questions about the nature of prosocial behaviour generally.

Chapter 2: Mere Categorization in Gains and Losses Contexts

In Chapter 2 I begin my empirical work by considering the two approaches that dominate psychological explanations of ingroup favouritism discussed above. As noted, one the one hand are scholars who focus on how simply sharing a group membership ('mere categorization') can induce ingroup favouritism, seemingly through social identity processes such as depersonalization leading to a transformation of motives and a desire to help ingroup members (Brewer, 1979; De Cremer & Van Vugt, 1999; Tajfel & Turner, 1979). On the other hand are those who advance a BGR model that focuses on expectations of reciprocity and reputational concerns (Yamagishi et al., 1999; Yamagishi & Kiyonari, 2000). A critical test of whether social identity can be sufficient to induce ingroup favouritism in minimal groups comes when decisions are private and so the participant cannot be identified as an ingroup member by the recipient. Unfortunately, existing evidence – using, for example, DGs – does not give a definitive answer.

In Chapter 2, I aimed to study whether mere categorization is sufficient to cause ingroup favouritism in minimal groups, while also considering three additional

questions: first, to what extent does economic efficiency influence the role of social identity and reputational concerns in ingroup favouritism; second, to what extent do social identity and reputational concerns represent complementary or competing processes within individuals; and third, to what extent ingroup favouritism might manifest differently in a *gains* context - helping another gain a good outcome – versus a *losses* context - helping another avoid a bad outcome.

Using a newly-developed measure – the Intergroup Lottery Task – I had participants decide whether they would pay some of their own money in the task to give another player a better chance of earning more money across a series of trials with different levels of economic cost and efficiency. I conducted this first in a gains context (Study 1) and then in a losses context (Study 2), and manipulated within-subjects across trials whether decisions were public or private, and whether the other player was an ingroup or outgroup member. My results showed that mere categorization was sufficient to induce ingroup favouritism: minimal groups were significantly more prosocial to ingroup members than outgroup members, and that this was the case regardless of whether decisions were public or private and regardless of the economic cost of helping. Moreover, the same pattern of results was observed across both gains and losses contexts.

In my first empirical chapter I demonstrate that social identity can be sufficient to induce ingroup favouritism. Particularly intriguing is that the effects of group membership were observed in both gains and losses frames and were evident regardless of the economic cost and benefit of helping. Such results suggest that ingroup favouritism might be a fairly automatic, intuitive process based on social heuristics rather than a

deliberative and controlled weighing up of the utility to be gained from helping an ingroup versus an outgroup member. Exploring this suggestion of ingroup favouritism as a kind of intuitive, automatic process was the focus of my next chapter.

Chapter 3: Intuitive Ingroup Favouritism

In Chapter 2, I found preliminary evidence that is consistent with ingroup favouritism being a relatively intuitive and automatic process. Indeed, there is evidence both that prosocial behaviour in general and that the preference for 'us' over 'them' is at least partially rooted in implicit, unconscious, or automatic processes. But to what extent is ingroup favouring prosocial behaviour rooted in such intuitive processes? Fairly little work has examined this, and this has been plagued by methodological problems that have led some to conclude that observations of increased prosocial behaviour under time pressure are merely an experimenter artefact, even outside an intergroup context.

In Chapter 3 I therefore investigated whether, when addressing the methodological problems in the dominant design, people under time pressure still act more prosocially than those under time delay and whether this is driven by time pressure increasing prosociality relative to an unconstrained control condition, and/or time delay decreasing prosociality relative to a control. And more importantly for the purposes of my thesis, I considered intuitive ingroup-favouring prosociality and looked at whether ingroup favouritism is observed even without the potential for deliberation; whether participants under time pressure act more prosocially with everyone or just ingroup members; and whether there are differences in prosocial behaviour under time constraint and towards ingroup or outgroup members depending on whether decisions are framed

as gains or losses. My results show that even after eliminating participant sampling bias and non-comprehension, time delay decreases cooperation relative to a control, and time pressure marginally increases cooperation relative to a control, and this was observed even towards outgroups and even in a losses context.

The first main contribution of this chapter is demonstrating ingroup favouring prosocial behaviour even under time pressure, and even after addressing methodological problems in previous studies, thus providing evidence that ingroup favouritism has its roots in intuitive processes. Even without time to deliberate, people help ingroup members more than outgroup members. Moreover, that ingroup favouritism is observed even in an unconstrained control condition highlights that cooperating with the ingroup more than the outgroup is a 'default' stance in human cognition. While we may be intuitive cooperators, we are not intuitive egalitarians. The second main contribution is to provide the first test of intuitive intergroup prosociality in both gains and losses frames. My results show similar patterns of group membership and time pressure in both frames: cooperation was greater towards ingroup than out members, and greater under time pressure relative to instructions to deliberate. This demonstrates the stability of the tendency for ingroup cooperation and favouritism.

Chapter 4: Reciprocity and Religious Prosociality

The previous two chapters have looked at minimal groups, but these differ (intentionally) from real groups in several ways. In my third empirical chapter I wanted to explore ingroup-favouring prosocial behaviour in the context of real groups, and look at whether there are differences in patterns of ingroup favouritism in non-interdependent

contexts where there is no potential for reciprocity (DG) compared to contexts where decisions are interdependent and expectations of reciprocity are critical (TG). Specifically, I chose to compare Christians and atheists for two reasons. First, in the US, there is substantial intergroup conflict between these groups which makes it an appropriate setting with which to explore intergroup prosocial behaviour. Second, there are features of religious belief specifically that make it a particularly interesting context in which to study prosocial behaviour. To quote Allport's (1954) highly influential *The Nature of Prejudice*, "the role of religion is paradoxical. It makes prejudice and it unmakes prejudice" (p.444). While some experimental research has suggested that religious people are only parochially prosocial towards co-religionists, other work has suggested that religious individuals are more prosocial in general, and other work has suggested that they are not more prosocial at all.

I looked at how much Christian and atheist participants transferred in a DG and TG to either Christian and atheist recipients. In both the DG and TG, Christian participants were more prosocial than atheist participants, and in the DG atheist participants, but not Christians, showed ingroup favouritism. Overall, Christian participants transferred more both to Christian and atheist recipients, and this was driven by the frequency of thinking about religious ideas. Interestingly, self-reported negative attitudes on the part of Christians towards atheists did not translate into actual discrimination. Such results provide further converging evidence that social identity can be sufficient to induce ingroup favouritism in anonymous, non-interdependent contexts, because atheists showed ingroup favouritism in the DG. Moreover, that even in

intergroup contexts characterized by distrust, this does not always translate into a lack of cooperation. I found that religious participants held negative attitudes about atheists and still transferred more to them in the games.

In addition to furthering our understanding of intergroup prosocial behaviour, my results provide further nuance into our understanding of religious prosociality. That Christian participants transferred more money to both Christian and atheist recipients in both a DG and TG is consistent with the claim that religious prosociality is a generalized phenomenon, not solely extended to co-religionists. Moreover, my results suggest that this was explained entirely by how often people thought about religious issues, thus representing reflection on moral norms within religious traditions that decree how to live. It is possible that this reflection leads individuals to build habits of virtue, and so – like the results of Chapters 2 and 3 – could suggest that religiosity helps inculcate social heuristics to cooperate with others.

Chapter 2: Mere categorization in gains and losses frames

This chapter is an extended version of a paper published as:

Everett, J. A. C., Faber, N. S., & Crockett, M. J. (2015). The influence of social preferences and reputational concerns on intergroup prosocial behaviour in gains and losses contexts. *Royal Society Open Science*, *2*(12), 150546. https://doi.org/10.1098/rsos.150546

Abstract

Can mere categorization be enough to lead to ingroup favouritism in prosocial behaviour, or is the potential for reciprocity and reputation-building necessary? How important are strategic economic concerns about the costs and benefits? And is this different depending on whether the prosocial behaviour occurs in a gains or losses frame? I explore this and find that participants are more prosocial towards ingroup than outgroup members and more prosocial when decisions are public than private, but find no interaction between group membership and either anonymity of the decision or expected economic value of helping. Contra to the claims of the BGR model, reputational concerns are not necessary for minimal group members to exhibit ingroup favouritism in the context of both gains and losses – heuristics based on social identity can be sufficient.

Introduction

As I have already discussed in my first chapter, two approaches dominate psychological explanations of ingroup favouritism. On the one hand are scholars who focus on how simply sharing a group membership ('mere categorization') can induce ingroup favouritism, seemingly through social identity processes such as depersonalization (Tajfel & Turner, 1979; Turner et al., 1987) leading to a transformation of motives and a desire to help ingroup members (Brewer, 1979; De Cremer & Van Vugt, 1999; Simpson, 2006). On the other hand are those who advance a bounded generalized reciprocity (BGR) model that focuses on expectations of reciprocity and reputational concerns (Yamagishi et al., 1999; Yamagishi & Kiyonari, 2000). I will not repeat the

theoretical arguments and empirical evidence for and against these approaches here, except to note that it remains an open scientific question.

To what extent do social identity processes and reputational concerns cause group favouritism? From the perspective of the BGR model, the critical test of whether social identity can be sufficient to induce ingroup favouritism in minimal groups comes when decisions are private and so the participant cannot be identified as an ingroup member by the recipient. Unfortunately, existing evidence – using, for example, DGs – does not give a definitive answer. There is a small but important literature showing that ingroup favouritism is observed in DGs only when dictators can be identified as ingroup members by the recipient and not when decisions are private, suggesting that reputational concerns are necessary for the behavioural expression of favouritism (Balliet et al., 2014; Mifune et al., 2010; Yamagishi & Mifune, 2008). However, other studies show that ingroup favouritism is observed even in anonymous DGs where reputational concerns are eliminated, suggesting that reputational concerns may be sufficient, but not necessary, to lead to ingroup favouritism (Rand et al., 2009; Yamagishi & Mifune, 2009). And finally, some studies show that ingroup favouritism in DGs is observed under anonymity, but is considerably stronger when participants can be identified, suggesting that both social identity and reputational concerns play an important role (Güth et al., 2009; Ockenfels & Werner, 2014). Given this ambiguity, I attempt to meet the challenge that Balliet and colleagues give: "new research is needed to unravel whether ingroup favouritism in situations lacking interdependence emerges because of social identity concerns,

reputational concerns, or perhaps some combination of these two" (Balliet et al., 2014, p. 1572).

The typical design used in previous studies to test whether social identity is sufficient to induce ingroup favouritism has three key features: it is between-subjects, it uses the DG, and it uses a one-shot single decision. Typically, that is, participants make a single decision about whether to transfer money to either an ingroup or outgroup member, under conditions of either anonymity or common knowledge (e.g. Ockenfels & Werner, 2014; Yamagishi & Mifune, 2008). This typical design does have some key advantages that deserve discussion. First, the advantage of using one-shot DGs (where one player makes a unilateral decision about how to divide an amount of money with a second player) is that there is no interdependence of outcomes, and so because an individual's decisions are not dependent on the behaviour of the other player (as they are in formal economic games like the PD or TG), one can be more confident that behaviour is driven by some internal motive (say, a desire to help the other person, or a concern with one's reputation) and not the external, interdependent, features of the situation. In the TG, for example, transfers by the first mover could be due to a prosocial motive, but could also be because the interdependent nature of the game means that their eventual payoff is dependent upon the other player such that they reason it advantageous to act prosocially. When this interdependence is removed - like in the DG - decisions can (in theory, at least) be attributed to more internal, less directly self-interested motivations. And by using a between-subjects design where participants make only a single decision, researchers can better reduce demand characteristics. These are perfectly reasonable

designs with many good points. As with any method, however, there are both strengths and limitations, and the use of this typical design does mean that certain questions have not been addressed.

In this chapter I wanted to explore mere categorization while making some methodological changes that would allow me to address some unanswered questions in the existing literature. First, I wanted to explore how concerns about economic efficiency influence the role of social identity and reputational concerns in ingroup favouritism, which has been impossible to test thus far with the use of single trials with a fixed cost. Second, I wanted to explore how social identity-based and reputation-based helping are correlated within-subjects. Previous research using a between-subjects design has only been able to look at the differences in conditions based on whether decisions are public or private and whether the recipient is an ingroup or outgroup member. I wanted to extend this by using a within-subjects design to see whether ingroup favouritism and general reputation-based helping in public conditions are correlated (as would be expected on the BGR) or not (as would be expected on social identity-based accounts). And third, I wanted to help fill an existing gap in the literature on ingroup favouritism by seeing whether framing the decisions as gains or losses matters.

Cost of Helping, Economic Efficiency, and Utility Calculations

The first change to the typical design I wanted to make was to vary levels of economic efficiency across multiple trials, rather than using a single decision with a fixed

amount of money (as in the DG). To do this, I applied a newly-developed task[24] used for studying prosocial behaviour at varying levels of efficiency to an intergroup context where decisions are either public or private, and call this the *Intergroup Lottery Task* (ILT). As described in more detail below, in this task participants decided whether they would pay some of their own money in the task to give another second player a better chance of earning more money. In this way, the task varies the economic cost to the participant of helping (e.g. £1 or £3), the probability of the other player getting a higher or lower payoff (e.g. 20% or 80%), and how much the other player would receive (e.g. £10 or £15).

Economic Costs and Ingroup Favouritism

Why might we think that economic considerations would influence observations of ingroup favouritism? Some scholars have suggested that observations of ingroup favouritism in minimal group contexts can be a result of participants dividing points or resources that have no financial consequences for the self (Yamagishi et al., 1999). Yamagishi and colleagues distinguish between ingroup favouritism – "allocation of reward of some utilitarian value in favour of ingroup members" – and non-material ingroup boasting – "more favourable evaluation given to ingroup members than outgroup members" (Yamagishi et al., 1999, p. 171). While there is evidence that mere categorization can lead to ingroup boasting, they argue, it does not always lead to ingroup favouritism because this requires perceptions of tangible ("utilitarian")

[24] The original lottery task was developed by my supervisor, Molly Crockett, during her own post-doctoral work. The original varied the economic cost to the participant of helping, the probability of the other player getting a higher or lower payoff, and how much the other player would receive. However, I adapted and expanded on this original design to create the Intergroup Lottery Task that included multiple trials in which decisions were public or private, and where the recipient was an ingroup or outgroup member.

interdependent outcomes and reciprocity. On the other hand, Balliet et al.'s meta-analysis shows that ingroup favouritism was observed for both hypothetical and real money. But this is not necessarily an either/or: there are various degrees of outcome distribution for the participant and recipient from those that have strong and tangible rewards to those that have more inconsequential and intangible rewards. On Yamagishi's reasoning one might expect ingroup-favouring allocations to occur primarily when it is not costly to the participant and has a large benefit (because this is more similar to ingroup boasting), but when such allocations are more costly to the participant, ingroup favouritism would disappear as reward-seeking considerations become more important.

Indeed, Klor and Shayo (2010) provide evidence that ingroup favouritism is sensitive to costs. They had participants in two natural groups randomly assigned gross incomes, and then had them vote over alternative redistributive tax policies. They found that social identity had a significant effect on preferences for the redistributive tax policies, with over one-third of participants voting for the policies that benefited their group, rather than their own monetary payoff. Participants tended to vote for high levels of redistribution when their group was overall relatively poor, even if they themselves were relatively rich. Despite this, however, decisions were affected by self-interest such that participants responded to the cost of helping the group: for every unit increase in the monetary cost to the participant, divided by the average benefit for the group, the probability of supporting the group-beneficial tax reduced by 8 percentage points (Klor & Shayo, 2010). In this way, individuals did not automatically support their ingroup, but only did this when the cost of doing so was not too high. Such findings are suggestive

evidence that sensitivity to costs and economic efficiency might moderate observations of ingroup favouritism.

Second, and at a more theoretical level, economic cost and efficiency would likely impact ingroup favouritism differently depending on whether ingroup favouritism functions more as a social heuristic (e.g. Rand, 2016) or deliberative cost-benefit utility calculation (e.g. Akerlof & Kranton, 2000). Some behavioural economists have argued that ingroup favouritism results from social preferences such that ingroup favouritism is motivated by an analysis that one will receive more pleasure (or utility) from helping the ingroup relative to the outgroup (Akerlof & Kranton, 2000). Because I varied the cost of helping, the benefit to the recipient, and the probability of the outcome, I could explore this by obtaining a measure of the net expected value (EV) of helping for each trial (i.e. the expected value of helping, divided by the cost of helping). To the extent that individuals help ingroup members more because they expect to get more utility from doing so, I reasoned that when the expected value of helping is high, this EV should predominate over group membership because the prosocial option is attractive either way. Similarly, when the net expected value of helping is small, group membership should not be very important because the prosocial choice is not attractive either way. What is interesting, however, is what would happen when net expected value of helping is moderate. Here, to the extent that individuals engage in a deliberative weighing up of utility, we might predict a non-linear interaction between group membership and expected EV such that ingroup favouritism is most prominent in an intermediate range in which it is attractive to give to an ingroup member, but not to an outgroup member. It

is possible, however, that such a weighing up of expected utility is not the proximate driver of ingroup favouritism. Instead, individuals could be adopting a social heuristic (e.g. Rand, 2016) to cooperate more with ingroup than outgroup members. and this is applied regardless of the specific costs of helping[25].

By varying costs to the participant and economic efficiency, I could explore whether economic concerns moderated the observation of ingroup favouritism and therefore provide suggestive evidence as to whether ingroup favouritism in general occurs more from a cost-benefit weighing up of utility or instead from social heuristics. But moreover, by varying the costs I also hoped to explore whether the psychological motivations of social identity and reputational concerns in driving ingroup favouritism are themselves sensitive to the cost of helping.

Economic Costs in Public and Private Decisions

Would economic cost and efficiency have different influences on ingroup favouritism in public or private conditions? This has not been discussed in the literature and there are no definitive predictions from either a social identity or reputation-based approach (which is partly why it is so interesting to explore). As outlined below, there are certain considerations that might suggest different roles of economic cost and efficiency in the different theories, but it is important to highlight that this is largely speculative: the theories themselves do not directly discuss the impact of these factors.

[25] In this experiment, at least. While the money involved for the time spent is fairly high for typical experiments at my university, it is plausible that if more substantial amounts of money were involved (e.g. in the hundreds of pounds), the use of heuristics would be reduced.

On a pure SIT account (Tajfel & Turner, 1979; Turner et al., 1987), one would not predict the cost of helping to play an important role. SIT focuses on the accentuation of differences such that people are not only concerned with helping the ingroup, but with improving the standing of the ingroup in comparison to the outgroup. To this extent, economic efficiency should not be that important because participants need to maintain positive self-esteem, and in the experimental context this can only be achieved by favouring the ingroup (Hogg & Abrams, 1988). As I argued above, though, SIT does not seem to be the best account of ingroup-favouritism and instead it is likely more fruitful to focus on work exploring social identification and the transformation of motives (Brewer, 1979; De Cremer & Van Vugt, 1999) and the BGR model (Yamagishi & Kiyonari, 2000).

What would social identification approaches predict about the moderating role of cost of helping? According to these approaches, social identity processes can lead to a transformation of motives whereby self-interest at the personal level is redefined at the collective level and the outcomes for self and others in the group become practically interchangeable (Brewer, 1979). This might suggest that economic cost and inefficiency would be more important for explaining prosocial behaviour extended towards outgroups than ingroups (similar to the empathy-based argument of Stürmer & Snyder,

2009)[26]. That is, given that – especially in minimal groups – participants typically have little desire to harm outgroups, it might be that the difference in ingroup favouritism will be most pronounced when it is costly and inefficient to help, because then the transformed motive for the ingroup is more prominent and cost/benefit considerations with regard to the outgroup members decreases. Thus, one might expect economic efficiency to have little impact on social identity-based ingroup prosociality, but more of an effect on prosociality extended towards outgroups.

In the BGR model, ingroup favouritism is caused by a strategic concern with one's reputation and, thus, expectations of future reciprocity (Yamagishi & Kiyonari, 2000). In this way, one would expect economic cost and efficiency to be important drivers of prosocial behaviour towards both ingroup and outgroups. The BGR model assumes that cooperating selectively with the ingroup is a kind of strategic investment in one's reputation (c.f. Semmann, Krambeck, & Milinski, 2004), and such strategic investments should only make sense when there is a reasonably good cost-to-benefit ratio[27]. Thus, one might expect that helping ingroup members – in addition to only being observed in

[26] Stürmer and Snyder (2009) argue that shared group identity should activate motivational processes like empathy that have their evolutionary origin in kin selection theory (Hamilton, 1964), while helping outgroup members should be driven more by motivational processes based on social exchange and reciprocity. Indeed, they find that empathy was a stronger predictor of helping for homosexual AIDS volunteers (for whom the client was an in-group member) than for heterosexual volunteers (for whom the client was an outgroup member) (Stürmer, Snyder, & Omoto, 2005). Further work by Stürmer and colleagues (2006) has conceptually replicated this finding, showing that empathy had a stronger impact on prosocial helping intentions when the target was categorised as an in-group member (similar cultural background) than when the target was an outgroup member (different cultural background).

[27] Another alternative is that, through costly signaling (Zahavi & Zahavi, 1999), individuals might be most likely to help an ingroup member precisely when it is inefficient because this signals that one is especially loyal to the group. I shall not discuss this here, however, because my focus is on proximate-level social psychological theories, and my results do not show such an effect anyway.

public conditions – would be very sensitive to economic efficiency. Overall, then, it remains unclear whether and why economic cost and efficiency might influence ingroup favouritism under public or private conditions and there are many competing predictions. Given this, my first aim in this project was to explore such effects.

Within-Subjects Analysis

Second (and relatedly), I wanted to use a within-subject designs to look at the relationship between reputation-based helping (i.e. helping more in public than private) and ingroup favouritism (i.e. helping ingroup members more than outgroup members) within participants. In part this was a consequence of the structure of the ILT and wanting to vary economic efficiency: In the ILT, participants completed 10 trials in each condition (ingroup/outgroup vs private/public) with varying levels of the cost to the participant of helping, the probability for the recipient, and how much the recipient stood to benefit. In a between-subjects design where economic efficiency across the three parameters was also manipulated across participants, I would have needed to recruit an impossibly large sample. Luckily, this challenge presented an opportunity. For example, there are general benefits of within-subjects designs that I could take advantage of, such as reducing variance and controlling for individual differences across populations (Charness, Gneezy, & Kuhn, 2012). But more importantly, my within-subjects task allowed me to look at the relationship between both overall helping (e.g. whether people who act prosocially in one condition act more prosocially in the others) and differential helping (e.g. whether people that show ingroup favouritism are also particularly sensitive to

whether decisions were public and private). The between-subjects design used previously has made this impossible to test.

How would we expect these to be related? According to the BGR, ingroup favouritism results from expectations that ingroup members will act prosocially towards you, and this is why it should not be observed in private conditions. Taking this assumption, one would therefore predict ingroup favouritism to be positively correlated with reputation-based helping (i.e. helping more in public conditions than private) because they are both results of the same, or very similar, psychological process based on expectations of reciprocity and reputation-building. In contrast, to the extent that mere categorization is enough to induce ingroup favouritism through social identity processes, ingroup-favouring prosocial behaviour would have no correlation with general helping more in public than private conditions, because ingroup favouritism and reputation-based helping are the results of different psychological processes.

Gains and Losses

Third and finally, I wanted to explore how ingroup favouritism might manifest differently in a *gains* frame - helping another gain a good outcome – versus a *losses* frame - helping another avoid a bad outcome. As discussed in my thesis introduction, work from prospect theory has documented loss aversion, such that a loss has a stronger influence on choices than a gain of the same magnitude (Kahneman & Tversky, 1979). And in the context of prosocial behaviour, it seems that because inflicting a loss is seen as more harmful and fairness-violating than withholding a gain, individuals are more likely to help another avoid experiencing a harmful outcome than they are to help

provide a positive outcome (Baron, 1995, 1996; Leliveld et al., 2009; Van Beest et al., 2005). With regards to economic efficiency (independent of group membership), we might expect that this would be more important for losses frames than gains frames because as De Dreu and McCusker (1997) note, there is some evidence that in losses contexts people pay more attention to decision alternatives (e.g. De Dreu et al., 1992; Dunegan, 1993) and so loss-framed individuals might engage in relatively more systematic processing, paying more attention to the costs and benefits.

Second, and more importantly, almost no work[28] has considered how gains and losses frames might influence intergroup prosocial behaviour in non-interdependent contexts. Why might we expect gains and losses frames to play a role in ingroup favouritism? One may reason that because people are more motivated by fairness in losses frames, they should show no preference for ingroup or outgroup members. However, I consider this hypothesis to be implausible because it assumes that greater fairness concerns in losses frames are equal with regards to ingroup and outgroup members. In fact, evidence suggests that fairness concerns are actually more salient in interactions with ingroup members (Mendoza et al., 2014; Shinada et al., 2004), implying

[28] In their discussion of loss aversion, Van Beest et al. (2005) consider work on the positive-negative asymmetry effect whereby people show ingroup favouritism in allocation of positive, but not negative, resources (e.g. Mummendey & Otten, 1998). The problem for my purposes, however, is that such work does not measure prosocial behaviour because there is no cost to the participant that has a corresponding benefit to the recipient. While this work is excellent at highlighting the difference in positive and negative allocations in intergroup contexts, it does not – and was not intended to – tell us much about loss aversion in intergroup prosocial behaviour. Correspondingly, in the 'key' papers on the positive-negative asymmetry effect, there is no mention of loss aversion or prospect theory, simply because this is not what they were studying (Mummendey et al., 1992, 2000; Otten & Mummendey, 1999).

that ingroup favouritism would occur in both losses and gains frames – and perhaps could even be stronger in losses frames, where fairness is more salient.

If ingroup favouritism is likely to manifest in both frames, a second question arises: to what extent is ingroup favouritism in losses frames driven by the same psychological mechanisms as in gains frames? Because losses are perceived as more harmful than gains, social identity processes to enhance the welfare of ingroup members could be more potent in driving ingroup favouritism in losses frames. Alternatively, reputational concerns could increase ingroup favouritism in losses frames because fairness principles are seen as more important in interactions with ingroup members, and so violating such fairness principles in public towards ingroup members is likely to be more damaging to one's reputation. Given the paucity of research on this, my third and final aim was to conduct explore ingroup favouritism using two variations of my design: one in a gains frame, and one in a losses frame.

Present Research

In this chapter I aimed to advance our understanding of intergroup prosocial behaviour by adapting a new task to the study of mere categorization - the Intergroup Lottery Task. This allowed me to explore the extent to which ingroup favouritism can be explained through mere categorization by looking at the degree to which subjects were willing to spend money to help ingroup and outgroup members when these decisions were public (thus activating reputational concerns) or private (thus, in principle, eliminating the potential for reputation building). By varying the costs and economic efficiency across trials, I could explore how sensitivity to costs relates to both observations

of ingroup favouritism and the processes by which this is posited to occur by. By using a within-subjects design, I could explore whether social identity processes and reputational concerns represent competing or complementary processes. And by conducting the study in both a gains (Study 1) and losses context (Study 2), I could explore how loss aversion influences intergroup prosocial behaviour.

To recap the key theoretical motivation for this study, there are different perspectives on whether ingroup favouritism would be observed for public or private decisions. According to the BGR, ingroup favouritism should be observed "only when the [actor's] behaviour is known to the ingroup recipient who identifies the dictator as a member of her own group" (Yamagishi & Mifune, 2008, p.16). In other words, ingroup favouritism should *not* be observed when decisions are private and so the recipient does not know the membership or decision of the actor. In contrast, to the extent that mere categorization can lead to ingroup favouritism through social identity processes, ingroup favouritism should be evident regardless of whether or not decisions are public (Brewer, 1979; De Cremer & Van Vugt, 1999). And while the BGR model assumes expectations of reciprocity and reputational concerns to be in opposition to social identity processes, these do not necessarily represent competing processes but could be complementary (Brewer, 2008). Here, one would expect ingroup favouritism is observed in private conditions, but is stronger in public conditions.

Method

Participants

80 participants studying at Oxford University were recruited for the two studies. 21 participants were excluded from data analysis if they guessed hypotheses and/or the occurrence of deception in the study, or if technical problems led to unusable data. Therefore, the final sample used for analyses consisted of 29 participants in Study 1 and 30 participants in Study 2 (Study 1, N = 29, 15 females, M_{age} = 23, SD = 4.45; Study 2, N = 29, 17 females, M_{age} = 21, SD = 1.89). A power analysis revealed that for each study, 24 participants were required to detect a moderate effect size (η_p^2 = .10) at the 5% level with 80% chance, and so both of my studies were sufficiently powered to detect a medium-sized effect[29].

Design

I report two parallel studies in this paper where I manipulated whether the recipient of the prosocial action was an ingroup or outgroup member, and whether the prosocial decisions were anonymous or not. Therefore, both studies had a 2 (Group: ingroup vs. outgroup) x 2 (Decision setting: public vs. private) within-subjects design where participants decided whether they would pay some of their own money in the task to give a second player a better chance of earning more money. As described in detail below, the prosocial action took place either in a gains frame (Study 1) or a losses frames

[29] This is less than the effect size that Balliet et al. (2014) offer, but we conceived of the study and begun data collection before the Balliet et al. paper was published, and so, given the weight of evidence for ingroup favouritism, assumed a medium effect size. As I will return to in the discussion, though, it would have been preferable to have more participants and this does stand as a limitation of these two studies.

(Study 2). Studies 1 and 2 were distinct studies and run at separate times, but I report them in parallel in the interests of conciseness because the design and analyses for both studies were identical.

Procedure

Participants were recruited in groups of four. They were shown to the testing room, given an information sheet about the study, and gave written consent to participate. Immediately thereafter, participants were asked to draw a small slip of paper from a hat but to refrain from looking at it until they were in their own private cubicle. The study consisted of three stages: a minimal group induction, a lottery task, and a debriefing questionnaire.

Minimal Group Induction

As in the work of Yamagishi & Mifune, I assigned participants to groups using a classic minimal group induction procedure based on preferences for the artists Klee or Kandinsky (Tajfel, 1970). Participants were told that they would be taking part in a study on art and social preferences, and that their first task would be to judge some artworks. In individual cubicles, participants were presented with 6 pairs of images and were required to indicate which of each pair they preferred. Participants were told that they would be categorized into one of two groups based on their art preferences. In actuality, all participants were categorized into 'Group A'. When participants received their group membership information, the first picture that they chose was displayed on the screen to increase the believability of the group categorizations. Participants were then asked to

read a description of their group and indicate two examples of how that description was relevant to them in their everyday lives.

Intergroup Lottery Task

After group categorization and still in their individual cubicles, participants were asked to read the slip of paper they had drawn from the hat at the beginning of the study. Participants were told that these slips of paper assigned them to different roles in the upcoming task. In actuality, the slips of paper were identical and all participants were told that they had been assigned to play in the role of 'Decider' and would interact with two of the other participants in that testing session, who had been assigned the role of 'Receiver'. Participants were told that they had a starting payment of £18, while the receivers would have a lower starting payment (£5 in Study 1; £15 in Study 2), and that their own starting payment was higher due to them making the decisions in the task. Participants were told that all three other participants were taking part in an unrelated study, but only two of the participants had the potential to receive money in the task.

The basic design of the task was as follows. Participants completed 40 trials where they made a choice between two lotteries that could win (Study 1) or lose (Study 2) money for a receiver. Both lotteries were always displayed on the same screen and differed in the probability of the receiver winning money, with 10 trials for each condition in a random order for each participant. The amount of money at stake was either £10 or £5, evenly split across trials. One of the lotteries was always better, in terms of having a higher expected value (EV – calculated by multiplying the size of the stake with the probability of the outcome), which meant the receiver would have a greater chance of

ending up with more money. For some trials, it was free to choose the better lottery. For other trials participants had to pay either £1 or £3 from their own starting payment of £18 to choose the better lottery. Participants were informed that 1 of the 40 trials would be randomly chosen at the end of the experiment and the decisions they chose for that round would be played and money distributed accordingly. Participants were instructed to therefore play each individual round as if it were the chosen lottery. To ensure that participants fully understood the lottery task, instructions were first presented in writing, then again orally by the experimenter, who ensured that all instructions and the task were understood and answered the participants' questions

In Study 1 – the gains frame - participants paid money for the other player to have a higher chance to gain more money (see **Figure 2.1**). In Study 2 – the losses frame – participants paid money for the other player to have a reduced chance of losing money (see **Figure 2.2**). Note that the probabilities, cost for the participants, and money for the other player was consistent across Studies 1 and 2, with the only difference being whether this choice was in the context of helping the other play gain money, or avoid losing money.

Figure 2.1: Example decision screen for the gains-frame lottery task (Study 1).

The figure shows that on a given trial, participants might have a choice between a left-hand lottery where they pay £0 from their own money to give another player a 20% chance of winning £10, or the right-hand lottery where they pay £1 and the other player has an 80% chance of winning £10.

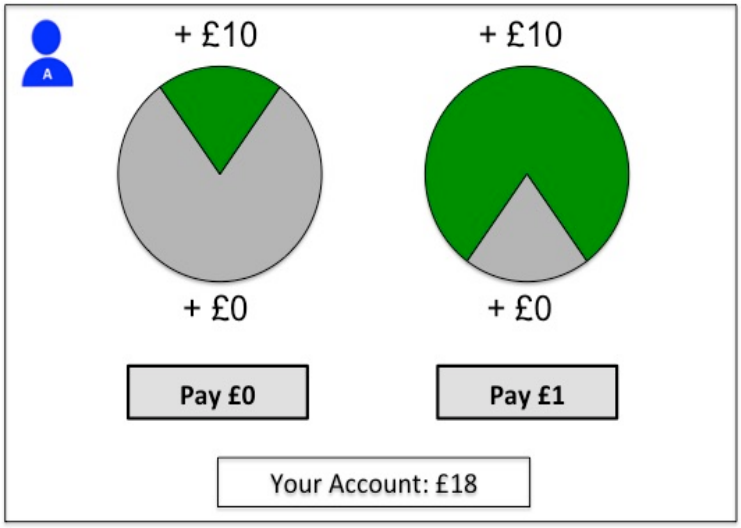
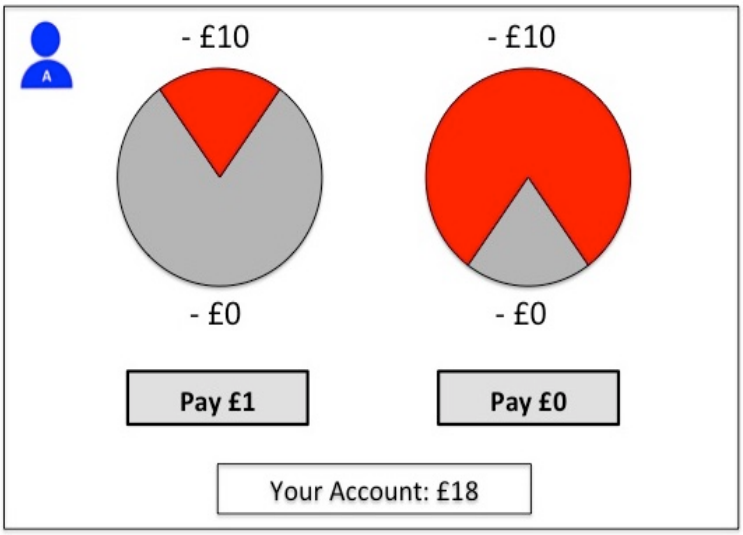


Figure 2.2: Example decision screen for the losses-frame lottery task (Study 2)

On a given trial, participants might have a choice between a left-hand lottery where they pay £1 from their own money to give another player only a 20% chance of losing £10, or the right-hand lottery where they pay nothing and the other player has an 80% chance of losing £10.



Participants made decisions in four experimental conditions derived from a 2 (Group: Ingroup vs. Outgroup) x 2 (Decision Setting: Public vs. Private) within-subjects design. To test whether the incidence of prosocial behaviour would be greater for ingroup or outgroup members, participants were told that they would play with two receivers: one from their own group (Group A), and one from the other group (Group B). In each trial, an icon in the top left-hand corner of the screen indicated whether the receiver for that round was an ingroup or an outgroup member.

To explore the role of social identity and reputational concerns, for half the trials participants were told that decisions were public (meaning that if that trial were selected at the end, the receiver would learn the decision and group membership of the

participant), and half the trials were private (meaning that the receiver would not find out about the outcome of the lottery - or even that the deciders had been playing a game that would determine their payment for the study). Public trials were intended to raise reputational concerns, whereby participants could be motivated by the desire to preserve a positive reputation, while conversely, private trials were decisions were anonymous were included to minimize such reputational concerns. For private trials, the icon in the left-hand corner had a blindfold, while for public trials there was no blindfold. To ensure anonymity for the private trials, participants were made aware that during the task the receivers would not know what the deciders were doing and were completing a different task in their own cubicles. Participants were told that receivers would only learn about the lottery task if a public trial was selected at the end of the task. Therefore, if a private trial was selected the recipient would not only be unaware of the participant's group membership, but also not know that the participant had been making decisions concerning the recipient's payoff at all. Participants were given a short series of questions to ensure they understood the task and the different trials, and a legend of these symbols was printed and placed next to the computer. In this way, we could ensure that participants understood the game before playing.

My dependent measure was a binary variable indicating whether participants chose the better lottery on each trial. I independently varied the probabilities of the lottery to win or lose money (e.g. 25%) the sum to be earned or lost by the receiver (e.g. £10), and the cost to the participant for helping (e.g. £3), which resulted in variation, across trials, of the net expected value (net EV) associated with choosing the helpful option. Net EV

was calculated by dividing the cost of choosing the better lottery from the gain in EV resulting from choosing the better lottery. For example, on trials with a low net EV, participants would have to pay a high cost (e.g., £3) to achieve a relatively small gain in EV (e.g., switching from a 10% chance of winning £5 to a 20% chance of winning £5). On trials with a high net EV, participants would have to pay a small cost (e.g., £1) to achieve a relatively large gain in EV (e.g., switching from a 10% chance of winning £10 to a 90% chance of winning £10). Excluding trials where the cost of choosing the better lottery was free, net EV took the form of three values: 1.00, 1.33, and 3.00. I then recoded these three values of net EV as three non-numeric predictors indicating low, medium, and high expected value. Coding this variable categorically, as opposed to continuously, allowed me to test for nonlinear as well as linear effects of net EV.

Prosocial behaviour, or helping, was defined as participants paying some of their own money in the ILT to give the receiver a better chance of winning (Study 1) or a lower chance of losing (Study 2), and rates of helping were averaged across trials for each of the four conditions.

Post-Task Questionnaire

After completing the ILT, participants completed a final debriefing questionnaire that included demographic information and questions assessing the extent to which they had understood the task and believed the deception. Participants were asked: "What do you think was the purpose of this study?"; "During this experiment, do you think we deceived you about anything?"; and "Have you heard of the Tajfel ('Klee Kandinsky') minimal group studies before?". These questions were presented in the order listed, and

on different pages so that the specific question on the Tajfel minimal group studies could not bias earlier responses on the first two questions. Participants who reported disbelieving the experimental instructions and paradigm were subsequently excluded, with the most common – and important – reason for exclusion being guessing that all participants were deciders, or that the groups were not real. After the experiment was finished it was made clear to participants that their decisions did not influence the payoffs of other participants, and that other participants would not know their decisions.

Analysis

In both studies I analysed the data using repeated-measures ANOVAs and the generalized estimating equation technique (GEE) (Hanley, Negassa, Edwardes, & Forrester, 2003; J. W. Hardin, 2005). The GEE procedure extends a generalized linear model to allow for analysis of repeated measurements. This technique allowed us to estimate a binary dependent variable (helping) in a fully within-subject factorial design. In this model, I specified two categorical predictors (group condition; decision setting), and the categorical predictor of low, medium, and high net EV (the extent to which choosing the better lottery increased the expected value of the receiver's outcome, minus the cost to the participant; categorized as low, medium, and high). The dependent measure of helping was a binary response, indicating whether participants chose the better lottery for the receiver or not.

Results

I first looked at rates of helping - the extent to which participants paid some of their own money to benefit the receiver - in a repeated-measures ANOVA (Group:

ingroup vs. outgroup; Decision Setting: public vs. private; see **Table 2.1** for *M*s and *SD*s). Overall, results revealed helping in about 60% of trials, with slightly higher overall rates of helping in the losses context (Study 2, $M = 0.64$, $SD = 0.28$) than in the gains context (Study 1, $M = 0.57$, $SD = 0.23$). In Study 1, there was significant main effect of group membership such that participants were more prosocial towards ingroup members than outgroup members ($F(1,28) = 6.36$, $p = .02$, $\eta_p^2 = .19$). In Study 2, the same pattern was observed, where this was marginally significant ($F(1,29) = 3.99$, $p = .055$, $\eta_p^2 = .12$). There was a main effect of decision setting in both studies, (Study 1, $F(1,28) = 7.74$, $p = .01$, $\eta_p^2 = .22$; Study 2, $F(1,29) = 16.89$, $p < .001$, $\eta_p^2 = .37$), such that participants were more prosocial in public decisions than private decisions (see **Figure 2.3** and **Figure 2.4**). Considering the relative effect sizes, it seemed that decision setting had a stronger effect on prosocial behaviour in the losses context ($\eta_p^2 = .37$) relative to the gains context ($\eta_p^2 = .22$), but that the effects of group membership were similar in both frames ($\eta_p^2 = .19$ vs. $\eta_p^2 = .12$). There was no significant interaction between group membership and decision setting in either study (Study 1, $F(1,28) = 0.40$, $p = .53$; Study 2, $F(1,29) = 0.02$, $p = .90$).

Figure 2.3: Helping as a function of group and decision context in Study 1

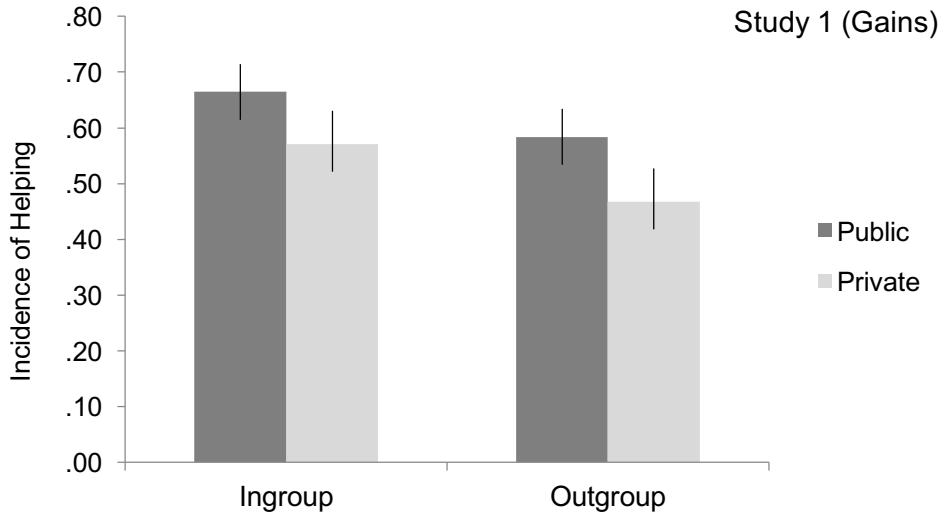


Figure 2.4: Helping as a function of group and decision context in Study 2

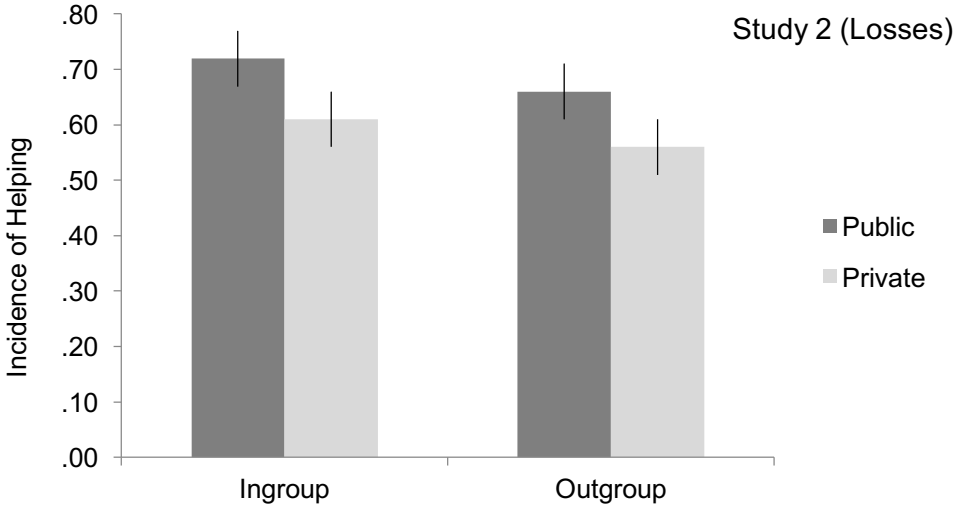


Table 2.1: Helping as a function of condition.

	Study 1 (Gains)				Study 2 (Losses)		
	Ingroup	Outgroup	Total		Ingroup	Outgroup	Total
Public	0.67 (0.26)	0.58 (0.30)	0.62 (0.25)	Public	0.71 (0.30)	0.66 (0.31)	0.69 (0.29)
Private	0.57 (0.30)	0.49 (0.26)	0.51 (0.26)	Private	0.61 (0.29)	0.56 (0.29)	0.59 (0.27)
Total	0.62 (0.26)	0.52 (0.25)		Total	0.66 (0.28)	0.61 (0.29)	

Note: SDs indicated in parentheses. Higher numbers indicate greater rates of helping.

I next ran a GEE analysis that included a regressor indicating the net EV, to test whether the expected value of helping moderated the influence of preferences or reputational concerns on ingroup favouritism. In particular, I was interested in how the net EV (the expected value of helping, divided by the cost of helping) was associated with ingroup favouritism. To the extent that ingroup favouritism is caused by a weighing up of utility to be derived, one might predict a non-linear interaction between group membership and expected utility. Specifically, ingroup favouritism might be most prominent in an intermediate range in which it is attractive to give to an ingroup member, but not to an outgroup member. In contrast, if social preferences more closely approximate social heuristics to cooperate more with ingroup members, then no such interaction should be observed because participants are likely to apply the social heuristic regardless of the specific costs of helping.

In this GEE analysis, my results from the ANOVA were replicated such that in both studies there were significant main effects of group membership, (Study 1, χ^2 = 7.66, p = .006; Study 2, χ^2 = 3.92, p = .05) and decision setting (Study 1, χ^2 = 8.39, p < .006; Study 2, χ^2 = 18.01, p < .001). I also looked at the effect of net EV on prosocial decisions and found that (as would be expected on a simple cost/benefit analysis), the more benefit the prosocial action had for the receiver, and the cheaper it was for the participant, the more likely it was that people helped (Study 1, χ^2 = 38.47, p < .001; Study 2, χ^2 = 27.63, p < .001). However, the expected value derived from helping did not moderate the role of social preferences or reputational concerns in ingroup favouritism, as net EV did not interact significantly with group membership or decision setting. Consistent with ingroup-favouring social preferences functioning as a kind of social heuristic, for both gains (**Figure 2.5**) and losses (**Figure 2.6**), participants were more prosocial to ingroup than outgroup members (and in public rather than private decisions), but there were also lower rates of helping as expected value decreased.

Figure 2.5: Helping as a function of group and economic efficiency in Study 1

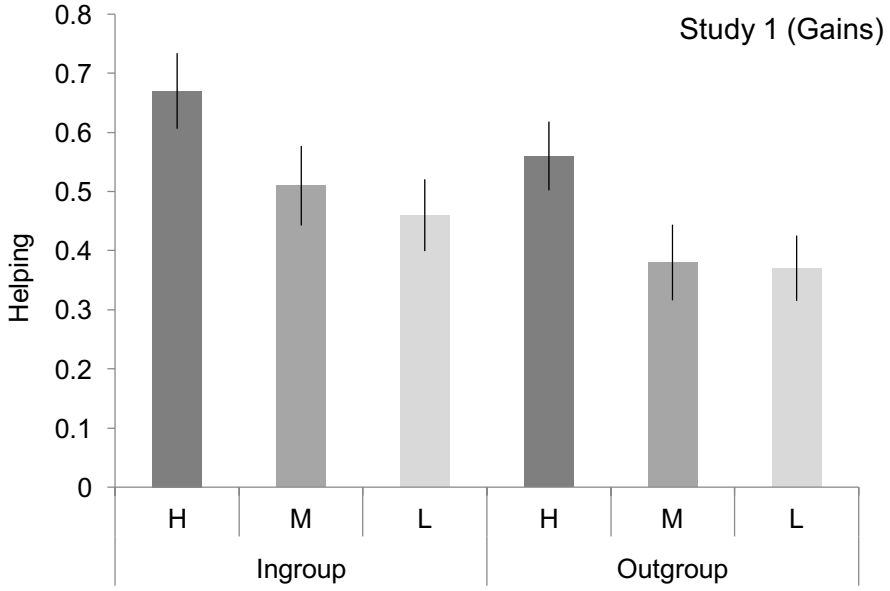
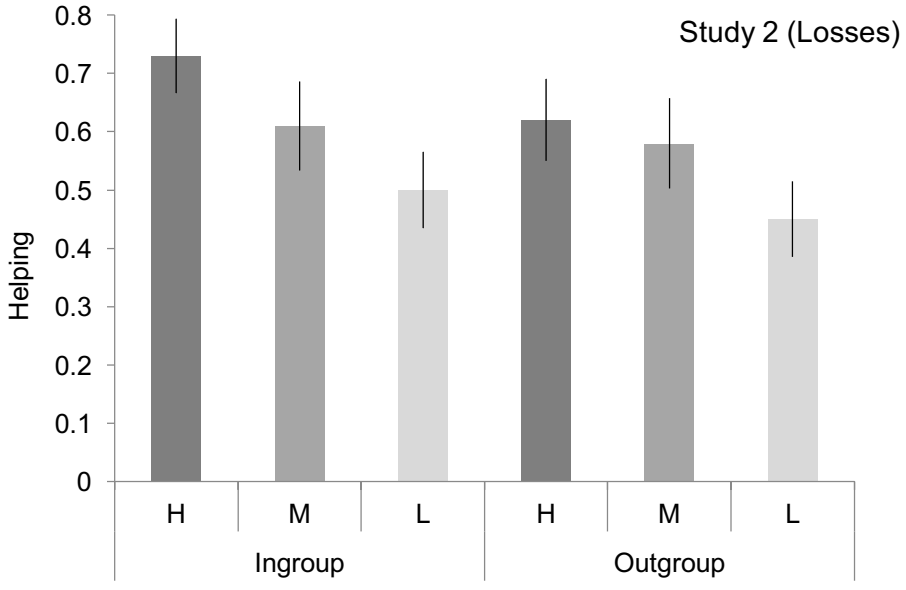


Figure 2.6: Helping as a function of group and economic efficiency in Study 2



Next, I looked at the relationship between helping in the different conditions. Across both studies, participants who helped in one condition were significantly more likely to help in the other conditions too (all *r*s > .51, all *p*s < .001), such that someone who helped an ingroup member more also helped an outgroup member more, and someone who helped more in public also helped more in private. This coheres with the idea from work on SVOs that people differ in their general tendency towards prosocial resource distribution (e.g. Messick & McClintock, 1968) and with other evidence on prosocial preferences being evident across different situations (e.g. Peysakhovich, Nowak, & Rand, 2014).

But to what extent were the differential rates of ingroup favouritism and helping in public more than private correlated? I found no correlation between ingroup favouritism (i.e. the rate of helping ingroup members, minus rate of helping outgroup members) and strategic helping (i.e. the rate of helping in public conditions, minus the rate of helping in private conditions), and this was the case both in Study 1 (r = -.26, p = .18) and Study 2 (r = -.16, p = .39). That is, the tendency to help ingroup members more than outgroup members and the tendency to help in public more than private were relatively distinct. In contrast, computing an ingroup favouritism score for both public trials and private separately revealed that ingroup favouritism in public trials was significantly correlated with ingroup favouritism in private trials, in both Study 1 (r = .51, p = .004) and Study 2 (r = .61, p < .001). This suggests that ingroup favouritism is driven by a motivational process that is fairly consistent across both gains and losses and public and private frames.

Discussion

In this chapter I applied a new experimental paradigm – the Intergroup Lottery Task - to explore the psychological mechanisms underlying ingroup favouritism. The central research question I addressed in this chapter concerned the extent to which mere categorization can lead to ingroup favouritism, or if instead the possibility for reciprocity and reputation is necessary. To test this, I looked at the incidence of prosocial behaviour towards ingroup and outgroup members in a minimal group paradigm under conditions of anonymity or not. The work I have reported in this chapter makes numerous contributions to studying ingroup favouritism in prosocial behaviour.

Contributions

First and foremost, I show that that mere categorization can be sufficient to induce ingroup favouritism in anonymous, non-interdependent contexts. According to the BGR model of Yamagishi and colleagues, ingroup favouritism should be observed only in public conditions where preservation of a positive reputation is possible (Mifune et al., 2010; Yamagishi & Kiyonari, 2000). In contrast, to the extent that social identity processes relating to a transformation of motives drives ingroup favouritism, this should be evident regardless of whether participants' decisions are seen by others (Brewer, 1979; De Cremer & Van Vugt, 1999). Across both studies I found that members of minimal groups were significantly more prosocial to ingroup members than outgroups, and that this was the case regardless of whether decisions were public or private. Thus, contra the claims of Yamagishi and colleagues, mere categorization into social groups can be sufficient to lead to ingroup favouritism, even when decisions are entirely private and there can be no

chance for reciprocity. As discussed in my first chapter, this is inconsistent with some of the previous literature (Balliet et al., 2014; Mifune et al., 2010), but converges with other papers that do report ingroup favouritism in private conditions (e.g. Güth et al., 2009; Ockenfels & Werner, 2014; Rand et al., 2009). Moreover, by using a within-subjects design I was able to show that participants' degree of ingroup favouritism (i.e. helping ingroup members more than outgroup members) is relatively distinct and uncorrelated with the degree to which they exhibit reputation-based helping (i.e. helping more in public than private conditions). Participants who showed ingroup favouritism in public conditions were also more likely to do so in private conditions, and this was consistent across both gains and losses frames. This is difficult to explain on a BGR model and - along with the result that ingroup favouritism was indeed observed in private conditions - is more consistent with ingroup favouritism being a fairly stable phenomenon that results from a different motivational process to more strategic, reputation-based helping. This, combined with my finding that ingroup favouritism was observed regardless of the economic cost and benefit of helping, is suggestive of ingroup favouritism being driven by a general social heuristic, rather than a deliberative and controlled weighing up of the utility to be gained from helping an ingroup versus an outgroup member. Overall, it seems that mere categorization is enough to induce ingroup favouritism, seemingly through activating social identity-based heuristics to help ingroups more than outgroups, and these heuristics seem to be applied regardless of the specific details of the context at hand (e.g. whether it is gains vs. losses; public vs. private; efficient vs. inefficient). Of course, this does not mean that expectations of reciprocity and reputation-building are

unimportant: to the contrary, the fact that prosocial behaviour was greater in public than private conditions shows it is a critical feature. But what these results do show, however, is that social identity processes can be sufficient to cause ingroup favouritism in minimal groups lacking interdependence. Ingroup favouritism cannot be explained solely through expectations of reciprocity[30].

The second main contribution of my work here is to provide one of the first examinations of ingroup-favouring prosocial behaviour in both gains and losses frames. While the majority of research on prosocial behaviour has used gains frames, everyday prosocial behaviour often takes the form of helping another avoid losing something. Given this, the extent to which ingroup favouritism would manifest and be driven by the same psychological mechanisms in both gains and losses frames is an important – and previously unaddressed - topic of study. I found the same pattern of results across both frames: ingroup favouritism was observed, and this was not affected by whether a decision was public or private. Thus, while my results show no evidence of gains or losses frames having markedly different effects on intergroup prosocial behaviour, I did find that whether a decision was public or private had a stronger influence – in terms of effect sizes – on prosocial behaviour in a losses context. Independent of ingroup favouritism, people are more likely to help when decisions are public rather than private, but this is especially so when the prosocial action is to help another avoid losing something. This is consistent with work on the do-no-harm principle (Baron, 1995, 1996) such that norms of

[30] At the proximate level, at least. Whether direct and indirect reciprocity, at an ultimate level, can explain all of prosocial behaviour is a different question and not the one I address here.

fairness may be more salient in loss frames and so the costs of publicly violating these norms are stronger. I continue to explore the effects of gains and losses frames on intergroup prosocial behaviour in my next chapter.

Limitations

There are several limitations of this work that deserve to be highlighted. One issue is sample size. Both studies had a final sample of 29 participants and, while an a priori power analysis revealed that 29 participants per study would have sufficient power, this is still a fairly low number. There were multiple reasons for this, including: the small active participant pool at Oxford; the difficulty of recruiting four participants to be in the laboratory at the same time; the expense of the study; and a larger than expected number of participants needing to be excluded. With more power, it is possible that the interaction between gains or losses frame, anonymity of decisions and/or group membership would have been significant. On the other hand, it is important not to overstate this limitation. The critical test was whether ingroup favouritism would be observed in private conditions, because this is the central point of contrast between explanations based on social identity processes and the BGR model. And, indeed, this main effect was significant in both studies.

Another potential concern is that I conducted two separate studies for the gains and losses framing. This was a deliberate choice because we judged that it would be too complicated for participants to answer both gains and losses trials, given that they already had to attend to multiple features of each decision (i.e. how much the decision cost, what the probabilities were, what amount the recipient stood to profit, whether it

was public or private, and whether the recipient was an ingroup or outgroup). To introduce gains and loss as an additional factor would likely have been too confusing. That said, in hindsight it would have been preferable not to run them as separate studies but instead have gains and loss as a between-subjects factor and the rest as within-subjects. This would have enabled me to look at the statistical interaction of gains vs. loss with the varied factors, though would have come with the corresponding issue of power, since to reliably detect this further interaction I would have needed a larger sample size. It was through recognising this limitation that I ensured the study reported in Chapter 3, which also looked at gains vs. losses, manipulated this as a between-subjects factor.

To what extent were reputational concerns fully eliminated? It is, of course, impossible to say. That said, every effort was made to make it clear to participants that if a private trial was chosen at the end as the 'real' trial, these decisions would stay private. We emphasised to participants that two of the three other participants were completing a different task, and that if a private trial was chosen at the end, they would not even know the 'real' participant was making decisions that affected their payout. We ensured that participants who disbelieved the experimental set-up were excluded from the analyses. From the participant's perspective, it was impossible for the experimenter to know when paying them at the end of the study whether they had acted prosocially or not because the amounts differed across trials and there was the probabilistic lottery. And even if there were vestigial implicit worries that private decisions were not private, because there were four people in each session and told participants there were two participants playing the 'real' game and two playing the other task, even if one of the

recipient-participants did think that one of the real-participants was making decisions involving their payoff, it would have been impossible to know who it was. While reputational concerns could not be entirely eliminated, therefore, it is reasonable to assume that participants were treating the private trials as exactly that - private.

Conclusion

In conclusion, in this chapter I looked at whether mere categorization can be sufficient to induce ingroup favouritism in both gains and losses contexts, drawing on both classic intergroup research and more recent advances in decision making. My results are consistent with ingroup favouritism resulting from more intuitive social heuristics than a strategic and deliberative weighing up of the value of helping. In my next chapter, I wanted to take a complementary approach and again look at ingroup favouritism in minimal groups in both gains and losses contexts, but this time focusing on the use of time pressure manipulations to encourage intuitive or deliberative thinking.

Chapter 3: Intuitive Ingroup Favouritism

This chapter is an extended and revised version of a manuscript published as:

Everett, J.A.C., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behavior–even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *Journal of Experimental Social Psychology*. 73, 76-81. https://doi.org/10.1016/j.jesp.2017.06.014

Abstract

By many accounts prosocial behaviour is a default strategy in social interaction. There are, however, several documented instances in which intuitive responding favours ingroup favouritism. I conduct a rigorous test of potential boundary conditions of intuitive prosociality by looking at whether intuition favours cooperation even towards competitive outgroup members, and even in losses contexts. Moreover, I address three major methodological limitations of previous research in this area: a lack of an unconstrained control condition; high levels of non-compliance with time manipulations; and non-comprehension of the structure of the game. Even after eliminating participant sampling bias and non-comprehension, I find that intuition does favour cooperation: even in competitive contexts towards outgroups and even in a losses context. Critically, higher cooperation with ingroup than outgroup players is consistent across conditions. People may be intuitive cooperators, but they are not intuitive egalitarians.

Introduction

In the previous chapter I provided evidence that prosocial behaviour in general, and ingroup-favouring prosocial behaviour specifically, might be driven more by social heuristics than a deliberative weighing up of the value to be obtained from helping. That prosocial behaviour is driven more by social heuristics is consistent a recent body of research suggesting that far from requiring effortful control, behaving prosocially arises from processes that are more intuitive and automatic (Zaki & Mitchell, 2013). For example, several studies find that people tend to make prosocial decisions in economic

games more quickly than selfish ones, and time-pressure increases the incidence of prosocial behavior (Rand, 2016; Rand et al., 2014; Rand, Greene, & Nowak, 2012). Time delay reduces helping in a 'dropped-glove' field study (Artavia-Mora, Bedi, & Rieger, 2017) and even risking one's own life to save another seems driven primarily by intuitive processes (Rand & Epstein, 2014). To explain this, the *Social Heuristics Hypothesis* (SHH; Rand, 2016; Rand et al., 2014) posits that the social strategies typically successful in daily life (e.g., cooperation) become automatized as default responses, and that deliberation can override these defaults to modify behavior. While meta-analytic work has provided strong support for the claim that manipulating reliance on intuition through time pressure encourages prosociality (Rand, 2016; Rand et al., 2014), a recent Registered Replication Report (RRR) by Bouwmeester and colleagues (2017) finds only mixed support.

Such findings on intuitive prosociality, however, seem difficult to square with other research highlighting intuitive intergroup bias such that that the preference for 'us' over 'them' is at least partially rooted in implicit, unconscious, or automatic processes. Fairly little work has examined how intuitive ingroup-favouring prosociality is, and this area has been plagued by methodological problems that have led some to conclude that observations of increased prosocial behaviour under time pressure are merely an experimenter artefact, even outside an intergroup context. In this chapter I consider some of numerous challenges to such work on intuitive prosociality. There are at least three major methodological concerns: the failure to include a control condition, the high levels of participant exclusions which can lead to selection bias, and inconsistency in the use of

comprehension checks. My first general aim in this chapter was to develop a procedure that would address these concerns. Second, and more importantly for my thesis, I explore whether even after addressing these methodological concerns, intuition favours ingroup favouritism in prosocial behaviour. Third, following on from the previous chapter, I conduct the first experimental investigation of intuitive ingroup-favouring prosocial behaviour in both gains and losses contexts.

Intuitive Prosociality

To explain observations of intuitive prosociality (e.g. Rand et al., 2012), Rand and colleagues have proposed the *social heuristics hypothesis* (SHH: Rand, 2016; Rand et al., 2014). The SSH posits that cooperation is shaped by behaviours that were successful in the past, and that it is particularly intuitive processes that are shaped by experience:

> "The behavior that is typically advantageous in daily-life social interaction (i.e., that maximizes payoffs in the long run) is automatized as a social heuristic, or a generalized default response. Deliberation, conversely, allows us to adjust to the *specific* social situation we are facing at any given time, overriding the intuitive response if that response is not actually payoff maximizing in the current setting."
>
> (Rand, 2016, p. 1193)

In the context of daily life, most important interactions (e.g., with co-workers, friends, and family) are repeated ones. Thus, because cooperation is non-zero-sum, cooperating can be, in the long-term, payoff-maximising and so in one's self-interest:

cooperating with others today can encourage others to cooperate with you in the future through the dynamics of direct and indirect reciprocity (Nowak & Sigmund, 1998; Trivers, 1971). Conversely, when interacting in settings where future consequences are *not* enough to incentivize cooperation (e.g., one-shot anonymous laboratory experiments), it is never in one's self-interest to cooperate and, therefore, deliberation should favour selfishness.

While there is often little explicit connection made the two, the SSH is strongly related to the principle of adaptation from interdependence theory (Kelley et al., 2003; Kelley & Thibaut, 1978; Thibaut & Kelley, 1959). Recall that according to the principle of adaptation (Van Lange, 2012; Van Lange & Rusbult, 2011), repeated social interaction experiences can yield adaptations that are reflected in relatively stable orientations (e.g. SVO) to adopt specific transformations. And indeed, people with more prosocial orientations do seem to have life experiences (e.g. secure attachment) that would shape such prosocial orientations (Van Lange, Otten, De Bruin, & Joireman, 1997). Where the SSH importantly extends this analysis is by highlighting that such adaptation has more influence on intuitive rather than deliberative processes.

In addition to results from behavioural studies, support for the SSH comes from mathematical modelling by Bear and Rand (2016). They looked at a mathematical model of dual-process agents playing a PD and found that of all the possible strategies, the dual-process strategy that performed best (and thus that would be favoured by evolution, social learning, and/or strategic decision-making) was one in which deliberation was used to change from cooperation to defection in one-shot anonymous settings, while

strategies in which deliberation was used to switch from defection to cooperation did not perform well and was disfavoured. It seems that while intuition can promote prosociality, according to this model at least, deliberation can only ever favour selfishness.

Most important is a recent meta-analysis in which Rand (2016) looked at the effects of intuition and deliberation on what he refers to as *pure* (i.e. altruistic) cooperation and *strategic* cooperation. Rand tested two key predictions of the SSH: first, that experimentally promoting intuition over deliberation should increase pure (i.e. altruistic, or non-strategic) cooperation "because in social dilemmas, intuition can favour either cooperation or non-cooperation (depending on the individual), whereas deliberation always favors non-cooperation"; and second, that experimentally promoting intuition over deliberation should have no overall effect on strategic cooperation "because both intuition and deliberation may favor either cooperation or non-cooperation, depending, respectively, on the individual's past experiences and his or her explicit expectations about what strategy will be payoff maximizing in the current context" (p. 1193). To test these two key predictions, Rand analysed 67 studies with 17,647 participants that had used either cognitive load, time constraints, ego depletion, or induction (priming) to measure cooperation in the PD, PGD, TG, or UG. These studies were then coded for pure cooperation or strategic cooperation. Pure cooperative decisions were those in which non-cooperation was payoff maximizing regardless of the actions of the other player or players (e.g. decisions in one-shot PD and PGDs, and player 2 decisions in the TG). Strategic cooperation decisions were those in which cooperation could be payoff maximising given certain choices by the other players (e.g. decisions in repeated PD and

PGDs, decisions by player 1 in TGs and UGs, and decisions by player 2 in repeated TGs). Consistent with predictions, results from the meta-analysis showed that there was 17% more pure cooperation in intuitive conditions than deliberate ones, but no effect for strategic cooperation. Such findings, Rand argues, support the claim of SSH that intuition favours typically advantageous behavior and deliberation favours behavior that is payoff maximizing in the current situation. In particular, these findings demonstrate that intuition only promotes cooperation in those contexts where the SSH predicts it should (pure cooperation) and not when it shouldn't (strategic cooperation), and this is explained by the SSH through the process of maximising long-term payoffs. Despite this evidence, however, there have been growing concerns about some of the methodological issues in these studies.

Methodological Challenges for Intuitive Prosociality

Participant Exclusion and Selection Bias

An important limitation with current studies using time pressure to study intuitive prosociality concerns high levels of participant exclusions leading to the problem of *selection bias* (Tinghög et al., 2013). The original studies have often excluded participants who failed to make their decisions within the response window. These exclusion rates are typically very high, but more importantly asymmetric across conditions. For example, in Rand, Greene, and Nowak (2012), 48% of participants failed to make their decisions under time pressure in Study 6, and 46% failed in Study 7 relative to 19% and 10% in the time delay conditions, respectively. While by some accounts these effects hold when including non-compliant participants (Rand et al., 2014; Rand, Greene, & Nowak, 2013),

these exclusion practices introduce the possibility of selection bias. Eliminating participants who are too slow from the time pressure condition and not from the time delay condition disrupts random assignment to condition, meaning that the observed difference in cooperation could be driven by systematic differences between the participants rather than the manipulation. And of course, as these exclusions become higher and higher (e.g. around 50%) so too does the likelihood that selection bias is seriously confounding results.

Indeed, researchers have begun to accumulate evidence that previously demonstrated observations of intuitive prosociality might be an experimental artefact resulting from selection bias. For example, Tinghög et al. (2013) try – and fail – to replicate these results, and therefore conclude the results of Rand et al. were just "an artefact of excluding the about 50% of subjects who failed to respond on time" (p.427). This analysis is consistent with the recent Registered Replication Report (RRR) (Bouwmeester et al., 2017) in which researchers from 21 separate groups across 12 countries conducted pre-registered replications of Study 7 from Rand et al. (2012). This RRR showed that two-thirds of participants failed to make decisions within the allotted time and that the effect of time pressure on cooperation was only present when excluding such individuals, and absent otherwise. This, they conclude, is "consistent with the presence of selection biases and the absence of a causal effect of time pressure on cooperation" [31] (Bouwmeester et al., 2017, p. 1). Given that the strength of any meta-analysis, including Rand's (2014, 2016), is

[31] Of course, this data is also consistent with the possibility that individuals who fail to conform to the time-pressure treatment therefore show no effect of that treatment.

dependent upon the methodological soundness of the constituent studies, this raises major concerns. Consequently, it is critical to develop ways of exploring the effects of time pressure and time delay in such a way that reduces high exclusions. Doing precisely this what I wanted to do in this chapter.

Comprehension Checks

A similar problem arises with the use of comprehension checks, where large numbers of participants fail to correctly answer comprehension questions about the structure of the game after they have played it. Across the studies reported by Rand et al. (2012), comprehension checks were implemented after the game had been played to avoid suggesting a deliberative mindset to participants prior to decision making. And indeed, as Rand and colleagues report in a supplementary study, participants who complete comprehension questions before making their decision choose to contribute significantly less than those who complete the comprehension questions afterward (Rand et al., 2012). However, in practice this means that participants may play the game while not understanding it. For example, while included in the data analysis, 32% of participants in Study 1 of Rand et al. (2015) and 29% of participants in Cone and Rand (2014) failed one or more comprehension checks. While the effects of time pressure were robust to controlling for whether people understood the game, this remains a potentially problematic aspect of the dominant methodological design.

Lack of Control Condition

A final – and much less discussed - limitation of existing studies using time pressure is that they have typically contrasted a manipulation intended to promote

intuition against a manipulation intended to promote reflection, but do not assess these manipulations against an unconstrained control condition[32] (e.g. Capraro, Jordan, & Rand, 2014; Cone & Rand, 2014; Rand et al., 2014; Verkoeijen & Bouwmeester, 2014). Given this, it is difficult to determine the extent to which intuition is increasing prosocial behaviour and/or the extent to which deliberation is reducing prosocial behaviour. Why is this a problem?

Without knowing the degree to which intuition increases cooperation or deliberation reduces cooperation, it is hard to know where the focus of theorising on intuitive prosociality should be and what the "default" stance is when people are not instructed either to go with their first hunch or to second-guess themselves. Work has typically referred to "intuitive prosociality", but without a comparison to a control condition, it could equally be the case that we exhibit "deliberate selfishness". To the extent that deliberation and intuition are simply (as often assumed) end-points on a continuum, this is not a major problem – just like saying conservatives exhibit high need for cognitive closure is functionally the same as saying liberals demonstrate higher openness (Jost, Glaser, Kruglanski, & Sulloway, 2003). Nonetheless, it is always important – where possible – to maintain precision, and unless one explicitly compares

[32] While the method I focus on here is the time pressure one because this has been the most widely used in studies of intuitive cooperation, it should be noted that other work looking at intuitive processing using alternative, non-time-pressure manipulations (e.g. the Stroop task) do compare intuition against a control condition (De Dreu, Dussel, & Velden, 2015; Ten Velden, Daughters, & De Dreu, 2017). On the other hand, these studies compare intuition against a control condition, but do not include a reflection condition. This, of course, leads to a similar problem: how to determine the effects of intuition and reflection against a neutral baseline?

intuition and deliberation against an uncontrolled control condition, this precision is hard to come by.

Moving Forward

In this chapter, then, I wanted to make three methodological improvements to the standard design of using time pressure to study intuitive prosociality. First, I wanted to find a way to reduce the extremely high exclusions that are typically unmatched across conditions, thus raising problems of selection bias. Second, I wanted to ensure that participants who were in the task actually understood the interdependent structure of the game. Third, I wanted to look at the relative effects of intuition (time pressure) and deliberation (time delay) against behaviour in a neutral, unconstrained condition.

As I shall outline more in the methods section, I achieved this by making three changes to the standard procedure. First, I drastically reduced exclusion rates due to the response window, achieving an exclusion rate of only 2% of participants enrolled in the study (18 participants out of 1158). I accomplished this by providing participants with extensive comprehension training prior to the task, using alternative (i.e., non-PD) payoff matrices. This prepared them to quickly absorb and respond to information presented briefly in the main task. Second, I equated the exclusion procedure across the time pressure and time delay conditions by subjecting participants in the time delay condition to an additional, subsequent time-pressure trial, and then excluding their data if they failed to meet the response deadline on this subsequent trial (9 excluded in the time pressure condition, 9 for the time-delay condition). This achieved exclusion rates that were both exceptionally low and comparable across conditions. Third and finally, I

included an unconstrained control condition in which participants had no instructions about how fast or slow they should respond and could answer at their own pace.

Intuitive Prosociality in Intergroup Contexts

Intuitive Intergroup Bias

This thesis concerns prosociality in intergroup contexts, and I was particularly interested in whether intuition favours prosociality only for ingroup members, or even towards outgroup members. Considering this question helps to clarify at what 'point' ingroup favouritism emerges: is it that people are intuitively cooperative to people in general and then deliberation introduces intergroup bias, or is it that even at the earliest stages of processing people are biased in favour of the ingroup?

The SSH posits that cooperative strategies that are typically advantageous in daily life with repeated trustworthy interaction partners become internalised as intuitions and get overgeneralized to less typical settings. Here, we might reason that social heuristics, precisely because they are overgeneralized, will be applied even towards outgroup members and it is deliberation that 'corrects' this process and causes increased intergroup bias. This is supported by work from Ma and colleagues (2015), who conducted a PGD with minimal group members and found that conceptual priming of intuition encouraged equivalent contributions to both ingroup and outgroup members, while it was the priming of reflection lead to pronounced ingroup favouritism (Ma et al., 2015). Here, it seems, reliance on intuition encourages prosociality regardless of group boundaries and can even eliminate the oft-founded difference in prosociality extended towards ingroup and outgroup members.

But at a theoretical level, the idea that intuition uniformly favours cooperation—even with outgroup members—is at odds with much past evidence illustrating implicit outgroup bias and distrust. Simply representing a social interaction as 'intergroup' undermines trust in others (Insko & Schopler, 1998; Pemberton et al., 1996; Wildschut et al., 2003), and the added perception of competitive relations between groups increases conflict (Campbell, 1965; Sherif, 1966). While we typically feel empathy at the plight of others, people may experience pleasure in response to outgroup members' adversities (*Schadenfreude*) and displeasure in response to their triumphs (*Glückschmerz*) (Cikara & Fiske, 2013; Leach, Spears, Branscombe, & Doosje, 2003; R. H. Smith, Powell, Combs, & Schurtz, 2009).

Indeed, several lines of evidence indicate that a general preference for 'us' over 'them' is at least partially rooted in implicit, unconscious, or automatic processes. At the behavioural level, intergroup bias often occurs outside the realm of conscious awareness and, for this reason, can be remarkably difficult to control or change permanently (Greenwald, McGhee, & Schwartz, 1998). For example, a recent test of 9 interventions that reduce implicit prejudice in the IAT found that none lasted more than one day (Lai et al., 2016). Moreover, some work seems to have shown intuitive aggression towards outgroups - for example, in the Shooter task where participants are required to "shoot" armed targets and to "not shoot" unarmed targets (Correll, Park, Judd, & Wittenbrink, 2002). Participants under time pressure are more likely to mistakenly shoot other-race targets than same-race targets (Correll et al., 2002) and are more likely to mistakenly shoot minimal outgroup members than ingroup members (Miller, Zielaskowski, & Plant, 2012).

As noted above, one marker of intuitive processes is that they appear early in developmental ontogeny, and evidence suggests that a preference for ingroup members does indeed appear early in development (for a review, see Dunham, Baron, & Banaji, 2008). For example, three-month olds prefer own-race faces (Kelly et al., 2005) and even young children are more likely to help someone that speaks their language (Kinzler, Dupoux, & Spelke, 2012). Like adults, children of just three years old show implicit intergroup bias by over-categorizing angry facial expressions as Black, indicating a negative association with that group (Hugenberg & Bodenhausen, 2004). Overall, such results are consistent with the idea that an implicit preference for the ingroup emerges as soon as an ingroup–outgroup contrast is acquired (Dunham et al., 2008).

And finally, neuroscience research has shown that group memberships appear to generate bias at the earliest components of perceptual and emotional processes (for a review see Cikara & Van Bavel, 2014). Recall that another marker of intuitive processes is that they tend to engage brain regions associated with value and reward processing, while deliberative processes engage brain regions associated with cognitive control and planning (Lieberman, 2007). In line with this, evidence suggests that shared social identity can lead individuals to place a higher intrinsic value on the welfare of ingroup member. For example, Hackel, Zaki, and Van Bavel (2017) conducted an fMRI experiment in which they examined vicarious reward responses of NYU and Columbia student participants to witnessing the monetary gains of ingroup and outgroup members. This allowed the researchers to look at the value attached to ingroup members independent of strategic concerns, since no decisions by the participant were actually made when

viewing these monetary gains. Their results showed that self-reported identification predicted differential ventral striatum responses when ingroup and outgroup members won money. Given that previous work has suggested that the ventral striatum is associated with reward processing, such findings suggest that social identification can actually shape the intrinsic value attached to ingroup members' outcomes (Hackel et al., 2017)

Overall, carving the world up into 'us' and 'them' appears to be almost automatic, and we are acutely and intuitively sensitive to cues marking group membership. Of course we do not always act in discriminatory ways, and it is plausible this is because in everyday life we typically have the opportunity (even if not the motivation) to deliberate and so can act in a more egalitarian way. Following this reasoning, one might expect that even under time pressure when forced to rely on intuition, individuals will help ingroup members, and do so more than outgroup members. When given the chance to deliberate, however, this bias may be attenuated because individuals can 'correct' their intuition and realise that cooperating in this context, regardless of group membership, is strategically disadvantageous.

Indeed, some research suggests that intuition favours ingroup favouritism rather than impartial cooperation. De Dreu and colleagues have conducted research on the neuropeptide oxytocin and, in this work, found evidence suggesting that a reliance on intuition encourages ingroup favouritism. Drawing on (cultural) group selection

models[33] (e.g. Choi & Bowles, 2007; Henrich, 2004; Sloan Wilson & Sober, 1994), De Dreu and colleagues reason that if individuals with psychological tendencies to be biased in favour of the group would have had increased fitness, this should be regulated by evolutionary old neurobiological systems like oxytocin. Indeed, oxytocin does seem to have an important relationship in modulating ingroup favouritism and outgroup derogation ("parochial altruism"). For example, intranasally administered oxytocin seems to increase both ingroup cooperation and defensive aggression against outgroups (De Dreu et al., 2010), and oxytocin is associated with greater empathy for ingroup members (De Dreu, Greer, Van Kleef, Shalvi, & Handgraaf, 2011). Applying this to intuitive prosociality, in recent work De Dreu and colleagues have considered the relationship between oxytocin, cognitive taxation, and ingroup favouritism. For example, using the IPD-MD they find that individuals who are cognitively taxed through a Stroop interference task subsequently show greater parochial altruism (De Dreu, Dussel, & Velden, 2015), suggesting that "the effects of oxytocin on decision-making in intergroup settings emerge partly because individuals given oxytocin act more spontaneously and invoke less cognitive control than those given placebo" (Ten Velden, Daughters, & De Dreu, 2017, p. 2). Testing this, they replicated their previous work and found that

[33] This evolutionary account is not critical to my thesis, which is focused on social psychological proximate mechanisms. Briefly, though, the idea is that if certain groups were more successful over extended periods of time because they cooperated more, then members of such groups would have been more reproductively successful than other groups and evolution would select for greater prosociality at the level of the group such that individuals would inherit dispositions to engage in costly prosociality for the group (Sloan Wilson, Van Vugt, & O'Gorman, 2008; Sloan Wilson & Wilson, 2007; Sober & Sloan Wilson, 1998; Traulsen & Nowak, 2006). Cultural group selection models expand this by positing culture as a separate evolutionary system that works in tandem with genetic evolution to shape behaviour (Henrich, 2004; Richerson et al., 2016).

participants who were given a placebo and placed under cognitive taxation showed higher ingroup cooperation. Moreover, those receiving oxytocin showed more ingroup cooperation regardless of cognitive taxation, suggesting that "oxytocin-induced ingroup cooperation is independent of cognitive taxation and, therefore, likely to be intuitive rather than consciously deliberated" (Ten Velden et al., 2017, p. 1). With regards to the focus here, however, the key result is that for those given a placebo, cognitive taxation increases ingroup favouritism.

A final possibility is that even if there is an intuitive bias in favour of the ingroup, that this is fairly encapsulated from social heuristic-based cooperation. For example, Rand, Newman, and Wurzbacher (2015) explored whether time pressure favoured cooperation in two contexts that were "hostile" toward cooperation: playing with an outgroup member, and playing in an explicitly competitive context. They used a PD (Study 1) and PGD (Study 2) with real political groups (Obama vs. Romney supporters) and found that a time pressure manipulation increased cooperation to both ingroup and outgroup members, but that the difference between groups was maintained relative to an enforced time delay condition. That is, time pressure made people more cooperative, but did not increase or decrease ingroup favouritism. This, they argue, is evidence that social context does not fundamentally change the dynamics of cooperation and that "social context effects may enter the cooperative decision making process at a very early stage and are already firmly in place before deliberation begins to undermine cooperative defaults" (Rand et al., 2015, p. 164). This is certainly not inconsistent with the evidence reported above by De Dreu and colleagues, and both show that ingroup favouritism

occurs even at the intuitive level[34] - but it does suggest that there might be two kinds of intuitive cooperation occurring: one that is more general (Rand, 2016) and one that is more parochial (De Dreu et al., 2015).

Overall, then, some studies suggest that intuition reduces ingroup favouritism (Ma et al., 2015), other studies suggest that intuition increases ingroup favouritism (De Dreu et al., 2015; Ten Velden et al., 2017), and other studies suggest that intuition increases prosocial behaviour but neither increases nor decreases ingroup favouritism, potentially because an intuitive ingroup preference already runs deep (Rand et al., 2015). How can this inconsistency be reconciled? At least some of the inconsistency might be due to the fact that these different studies have used different methods of priming intuition: the research showing that intuition increases ingroup favouritism has used Stroop-interference tasks (De Dreu et al., 2015; Ten Velden et al., 2017); the work showing intuition reduces ingroup favouritism has used conceptual priming of intuition (Ma et al., 2015); and the work showing that intuition increases prosociality but does not eliminate ingroup favouritism has used time pressure manipulations (Rand et al., 2015). Part of the reason for these different might be that these manipulations differ in strength – an issue I return to in my discussion section of this chapter.

---

[34] When discussing these results, for example, De Dreu et al. (2015) appear to see no contradiction between the findings, writing "Our finding that cognitive taxation increases contributions not only fits [the results of Rand et al.], but also clarifies that increased cooperation with someone sharing the same political preference is motivated by a desire to benefit the in-group more than a desire to hurt the out-group" (p.7).

As one of the more widely used measures[35] of studying intuition in prosocial behaviour, I wanted to focus on time pressure. As discussed in the previous section, it is possible that some of the mixed nature of these findings is due to methodological problems like selection bias or the lack of a control condition. For example, previous work using time pressure has contrasted between a manipulation intended to promote intuition against a manipulation intended to promote reflection (e.g. Capraro et al., 2014; Cone & Rand, 2014; Rand et al., 2014) but has not included a control condition. This makes it impossible to know what the 'default' mode of interaction is in intergroup prosocial encounters: when people are not instructed either to act on their first impulse or to second-guess themselves, do people cooperate more with the ingroup? Because this work has contrasted two active manipulations without including a control condition, it is impossible to know what the default is.

Intuitive Prosociality in Gains and Losses Contexts

The final primary contribution I wanted to provide in this chapter was to look at intuitive prosociality in both a gains and losses context. All the studies of intuitive intergroup prosociality involve social dilemmas framed in terms of gains, where participants must choose whether to help others gain, or avoid losing, something positive. But as discussed in the previous chapters, there are several reasons to think that framing prosocial decisions in terms of gains or losses may matter to intuitive prosocial behaviour. Some evidence suggests that in losses contexts people pay more attention to

[35] As indicated, for example, by Rand's (2016) meta-analysis: there are 31 time constraint manipulations, but only 13 using conceptual priming and 8 using cognitive load.

decision alternatives (e.g. De Dreu et al., 1992; Dunegan, 1993), consistent with loss aversion relying on systematic processing. However, other more recent work has posited framing effects to result from more intuitive processes (Kahneman & Frederick, 2007), consistent with findings that loss aversion is associated with affective systems in the brain (Gonzalez et al., 2005) and is observed even under cognitive load (Whitney et al., 2008). To my knowledge, however, no studies have looked at framing effects in time pressure, and no work has considered the role of intuition and deliberative in driving framing effects in interpersonal contexts.

More importantly for this thesis, when it comes to loss aversion in prosocial behaviour it remains unclear whether framing social dilemma decisions as losses increases cooperation (e.g. Experiment 3 in Komorita & Carnevale, 1992), reduces cooperation (e.g. Brewer & Kramer, 1986; McCusker & Carnevale, 1995), or has no effect (e.g. de Heus, Hoogervorst, & Dijk, 2010). It seems that this is because decision frames have divergent effects based on an individual's prior motives such that prosocial people become more cooperative, and 'individualists' less cooperative (De Dreu & McCusker, 1997). Because people's prior motives are different with regards to ingroup and outgroup members, we might therefore expect loss frames to encourage defection towards outgroups (because without a prior motive to be cooperative, loss frames make people more selfish) and cooperation towards ingroups (because loss frames enhance the existing motive to help the ingroup member). However, the effects of decision frame in intuitive cooperation in general, and intergroup intuitive cooperation specifically, remain unknown, simply because there is so little work looking at this.

In the previous chapter I looked at how people made prosocial decisions in a DG-like task lacking interdependence and found little evidence of gains and losses having markedly different effects. As noted in the discussion, however, the robustness of these results is not without question and so I did not want to draw too strong a conclusion about the effects of gains and losses in general. First, given that I ran the studies separately at different points it was not appropriate to combine the studies together, and even if it were, the sample size is likely to be somewhat underpowered for directly comparing the strength of effects of gains and losses on intergroup behaviour. Second, it is also possible that gains and loss frames have a differential directional or strength of effect in games with interdependence. It is possible, for example, that in interdependent contexts a kind of second-order loss aversion phenomenon occurs whereby people anticipate that others will be more loss averse and less likely to cooperate, and consequently are less likely to cooperate themselves because they anticipate defection on the part of the other. In decisions lacking interdependence, such as in Chapter 2, this would not emerge. Third, no research has looked at prosocial behaviour framed as gains and losses in time pressure specifically, and there are reasons to think this might be important. For these reasons, in this chapter I wanted to look again at gains and losses, but this time using PDs and manipulations designed to induce promote intuition or deliberation.

Present Research

Method

Pre-Registration

My design, hypotheses, and analysis plan were all pre-registered at the Open Science Framework (https://osf.io/v82gc/) and all data, analysis code, and experimental materials are available for download at: https://osf.io/pz4he/.

Participants

1316 American participants completed the survey online using Amazon Mechanical Turk and received $0.70 for their time, with all participants being paid the same additional bonus of $0.70 (the maximum achievable in the task). Participants were excluded if they completed the survey more than once (N = 158); failed simple comprehension checks regarding to which group they had been assigned to (N = 0); and/or failed to the complete the main PD trial in the time pressure condition (N = 9) or the time pressure bonus round in the time delay condition (N = 9). Therefore, the final sample consisted of 1140 participants (559 female). Because final analyses in the paper looked only at the additive (main) effects, this translated to approximately 380 per condition for the three time-instruction conditions, and 570 per condition for the gains/loss and ingroup/outgroup conditions. While not the eventual focus of the analysis, I had nearly 100 participants for each cell of the 3x2x2 design, which I hoped would be sufficient to detect a moderate effect size.

Design

I used a fully between-subjects design where participants completed a Prisoner's Dilemma (PD: Axelrod, 1980; Rapoport & Chammah, 1965) and manipulated (1) the group membership of the other player; (2) whether participants were encouraged to make their decision intuitively, through deliberation, or in the absence of either instruction; and (3) whether the PD was framed in terms of gains or losses. This study had a fully between-subjects design with 12 conditions: 2 (Group Membership of Other Player: Ingroup vs. Outgroup) x 3 (Instructions: Intuitive vs. Deliberative vs. Control) x 2 (Frame: Gains vs. Losses).

Upon beginning the study, participants were told that they would be randomly assigned to one of two teams (Eagles or Rattlers) based on their responses to five short personality questions. Upon being assigned to a team (always the Eagles), participants were told these teams were competing with other in a problem solving task (Cikara, Bruneau, Van Bavel & Saxe, 2014). To simulate competition, participants were told that "members of the Eagles team have completed 82 tasks; members of the Rattlers team have completed 84 tasks. Whichever team successfully completes 100 tasks first will win extra cash. This is a competition between teams: Only one team will be awarded the bonus!" Participants were then told that before completing the competitive problem-solving task they would first engage in a short and entirely separate game (the PD) with another participant, where this other participant could either be from their team (ingroup), or the other team (outgroup). Like the original Tajfel studies, I wanted to use minimal groups to control for the effects of stereotypes and familiarity that can exist for 'real groups',

while making them competitive to simulate the functional relations between groups in conflict (L. W. Chang, Krosch, & Cikara, 2016).

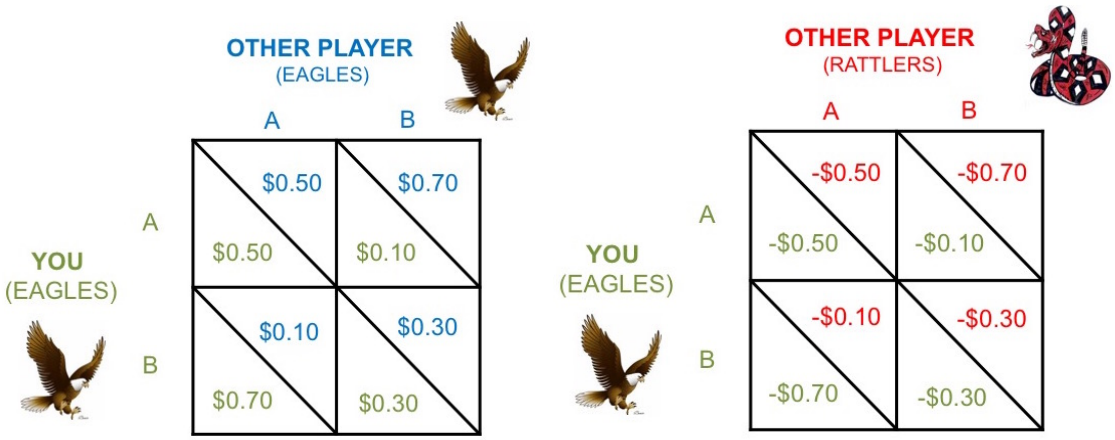
Participants were then given comprehensive instructions and extensive training on the structure of the PD to ensure that they understood the different possible outcomes of the game. All participants were required to manually enter the different outcomes that each player would receive depending on the choices of the other player in an example matrix, and participants had to enter these correctly before being able to move forward in the study. Participants could spend as long as they wanted on this section, and if they entered an incorrect answer were simply prompted to correct it in the appropriate place. Therefore, and in contrast to most previous studies, I did not have to exclude a single participant post-hoc based on failing to understand the structure of the game[36].

The PD was framed in one of two different ways. In the first, standard, version of the game – the gains frame - participants learned that they could earn an additional bonus of up to $0.70 depending on the decisions that both they and the other player made in the game. For example, if both participants chose to cooperate (A/A in the left panel of **Figure 3.1**), they would both gain $0.50 and so would end the game with a bonus of $0.50.

[36] This is not to say that everyone who opened the survey completed it. While 1687 participants opened the study link, only 1626 moved beyond the first page of the personality questions for the team assignment (dropout of 4%) and only 1316 moved beyond the instructions of the game and comprehension checks (dropout of 18%). This led to a total dropout rate of 22%, but one that was equivalent across conditions. As can be seen in Table 3.1, there were roughly the same number of people in both the in-group (N = 573) and outgroup (N = 567) conditions, and both the time delay (N = 366) and time pressure (N = 366) conditions. There were slightly more participants in the control condition (N = 407) compared to the time delay and time pressure, and slightly more participants in the gains conditions (N = 600) than the losses conditions (N = 540), but these differences were small considering the large sample size.

In the other frame – the losses frame – participants were told they had been given a bonus of $0.80 and that they would lose a certain amount (at least $0.10) based on the decision they made and the decision the other player made. For example, if both participants chose to cooperate (B/B in the right panel in **Figure 3.1**), they would both lose $0.30 and would end the game with a bonus of $0.50. Therefore, in both versions of the game, participant's choice to cooperate or defect would yield the same eventual bonus – but whether this was framed as gaining or losing money differed.

Figure 3.1. Example Matrices given to participants (left: gains frame with ingroup partner; right: losses frame with outgroup partner)



I encouraged participants to rely on intuition by requiring them to make their decision in less than 15 seconds (time pressure), encouraged to think deliberatively by requiring them to wait and think for at least 15 seconds before deciding (time delay), or maintained a neutral mind-set by giving them no instructions about how long to take

(control). This specific time to induce intuition vs. deliberation (15 seconds) was determined after running a pilot study (N = 102) in which, in the absence of any instructions to be fast or slow, participants took a median time of 17 seconds.

Participants in the intuition condition who did not respond within the allotted time were excluded from data analysis (N = 9). Furthermore, in order to ensure that participants in the time pressure and time delay conditions did not systematically differ in terms of their ability to answer in a short time-frame, participants in the time delay condition completed a bonus round of the PD (with an alternative matrix in which all values were reduced by 0.05) under time pressure after completing the main task. Participants in the time delay condition who failed to make a decision in the allotted time in this bonus round (N = 9) were also excluded from data analysis. This is notably lower than the exclusions typically observed in these studies, which can reach around 50% of the entire sample.

In the task, participants in the time delay condition took a median time of 23 seconds, and participants in the intuition condition took a median time of 9 seconds. Unfortunately, due to a technical error, response times for the control condition were not recorded.

Results

A logistic regression using *R* tested the effects of group membership, gains/losses frame, and time manipulation on the likelihood that participants made cooperative decisions in the PD. I looked at which model fit the data best by conducting a series of analyses that used likelihood ratios to compare the fit of the full summative model

(containing all predictors and their interactions) with various nested models[37]. I started by dropping the interaction terms and found that none of these added anything to the model. However, dropping any of the main effects (group; gains/losses frame; time manipulation) did result in significantly worse fit, showing that all three were necessary for analysis. Given that the additive model with only the main effects was more appropriate than a summative model with all interaction effects, this is the model I focused on for data analysis.

This model showed main effects of all three predictors (see **Table 3.1** and **Figure 3.2**). There was a significant effect of decision frame (B = -.42, SE = .12, Z = -3.52, $p < .001$) such that participants were more likely to cooperate under a gains frame (61%) than a losses frame (51%), and a significant effect of group membership (B = -.30, SE = .12, Z = -2.51, p = .01) such that participants were more likely to cooperate with an ingroup (60%) than an outgroup member (53%). Finally, participants were more likely to cooperate under time pressure (63%) than under time delay (49%) (B = .58, SE = 0.15, Z = 3.81, p = .004), and marginally more likely to cooperate when given no instructions (57%) than under time delay (B = .31, SE = 0.15, Z = 2.44, p = .08). Although participants cooperated slightly more under time pressure than when given no instructions, this difference was not significant (B = -.27, SE = 0.15, Z = -1.79, p = .17). Put simply, there were main effects of time pressure, group membership, and gains and losses frames, but no evidence of interactions between them.

[37] Thank you to my collaborator on this project, Zach Ingbretsen, for developing the code for the analysis to compare the models.

Figure 3.2: Cooperation in the PD as a function of gains/losses frame, group membership of the other player, and whether decisions were made under time pressure, time delay, or no timing instructions.

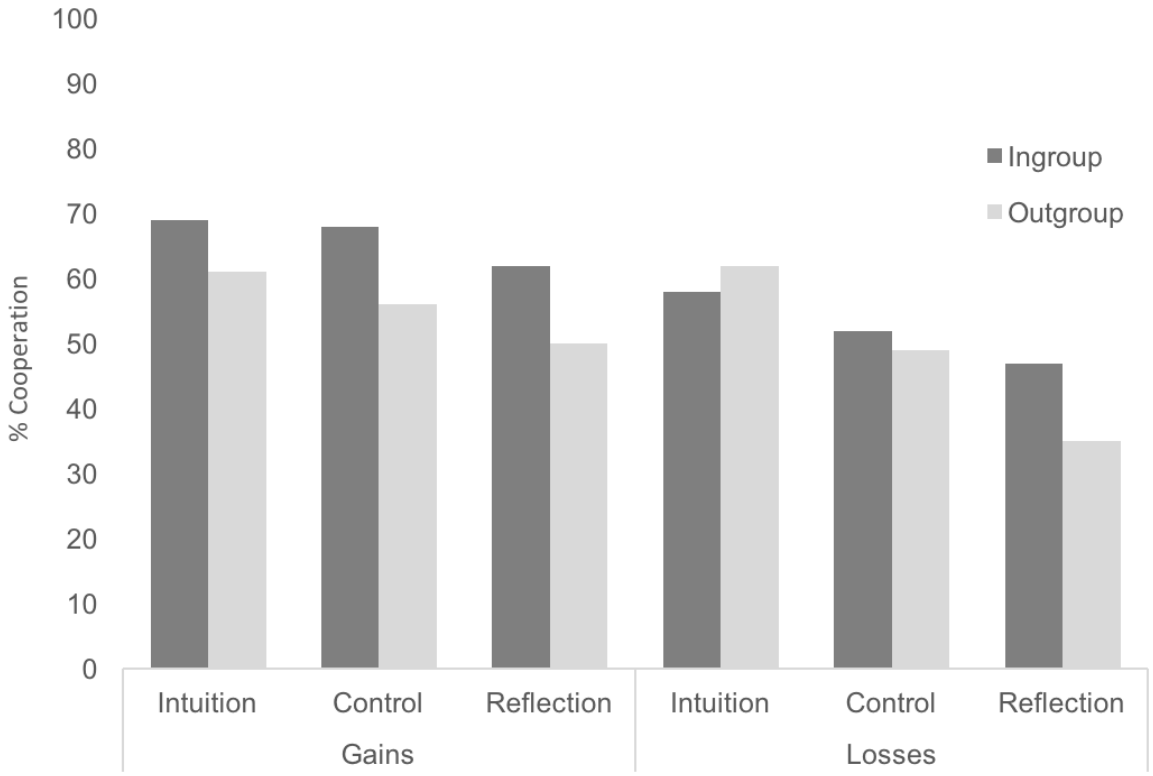


Table 3.1. Rates of cooperation across conditions.

	Gains Frame					
	Ingroup			Outgroup		
	Cell N	Freq. of Cooperation	Percentage Cooperating	Cell N	Freq. of Cooperation	Percentage Cooperating
Intuition	90	62	69%	99	60	61%
Reflection	96	60	62%	96	48	50%
Control	117	79	68%	102	57	56%
	Losses Frame					
	Ingroup			Outgroup		
	Cell N	Freq. of Cooperation	Percentage Cooperating	Cell N	Freq. of Cooperation	Percentage Cooperating
Intuition	84	49	58%	93	58	62%
Reflection	88	41	47%	86	30	35%
Control	98	51	52%	91	45	49%

Discussion

In this chapter, I explored intuitive ingroup favouritism and provided a rigorous test of potential boundary conditions of intuitive prosociality by addressing methodological concerns that have plagued previous studies. Even after addressing high

levels of participant exclusion and selection bias, I found that intuition favours cooperation even in competitive contexts towards outgroups, and even in a losses context, though the differences in cooperation between groups was consistent across conditions.

Contributions

Intuitive Ingroup Cooperation

The first main theoretical contribution of this work is to further our understanding of the relationship between intuitive processing and ingroup favouritism. That my results demonstrated ingroup favouring prosocial behaviour even under time pressure and even after addressing methodological problems, supports the idea that intergroup bias has its roots in intuitive processes. Even without time to deliberate and, presumably, consider strategic expectations about whether it is more advantageous to help ingroup members more than outgroups, people show ingroup favouritism. Moreover, that ingroup favouritism is observed even in an unconstrained control condition highlights that cooperating with the ingroup more than the outgroup is a 'default' stance in human cognition. Consistent with previous work highlighting the importance of intuitive processes in ingroup cooperation (Cikara & Van Bavel, 2014; De Dreu et al., 2015; Dunham et al., 2008; Rand et al., 2015; Ten Velden et al., 2017), my results highlight that while we may be intuitive cooperators, we are not intuitively impartial.

It is interesting that while I observed ingroup favouritism even when deliberation was impossible, time pressure did not seem to interact with group membership to make ingroup favouritism especially pronounced under intuition. This is also what Rand et al.

(2015) found using time pressure manipulations, but not what has been demonstrated using an IPD-MD and cognitive interference (De Dreu et al., 2015; Ten Velden et al., 2017). This does not change the core conclusion that ingroup cooperation and ingroup favouritism are supported by intuitive, automatic processes, but it does raise questions about why intuition did not increase ingroup favouritism. At least two potential causes stand out: while Rand et al. (2015) and I used time pressure and an online MTurk sample, De Dreu et al. used cognitive Stroop-interference tasks and had participants come into the lab. Given that Rand's (2016) meta-analysis suggests that the effects of intuition on cooperation are weaker when implemented through time pressure than when using alternative methods such as conceptual priming or cognitive load, it is possible that my manipulation was not strong enough to observe an interaction between group membership and time constraint. That said, my manipulation was strong enough to demonstrate the previously found difference in cooperation between time delay and time constraint. One possibility is that, as suggested in the introduction to this chapter, there are two kinds of intuitive processes going on. One is a general tendency for cooperation, and one is a tendency for ingroup favouritism, and it is possible that the general tendency is more easily manipulated than the ingroup favouring one. It would be interesting for future research to consider this more.

Gains and Losses

The second contribution of my work is to provide the first test of intuitive (ingroup) cooperation in both gains and losses frames. My results show similar patterns of group membership and time pressure in both contexts: cooperation was greater

towards ingroup than out members, and greater under time pressure relative to instructions to deliberate.

Time pressure had no greater effect on loss frames more than gains frames. This is consistent with work by Whitney et al. (2008) who show that working memory load does not impact the potency of framing effects, but is potentially surprising if one thinks that these framing effects can be overridden by the potential for deliberation (Kahneman & Frederick, 2007). Where this work extends on previous work, of course, is through showing this in prosocial behaviour rather than in money for the self or hypothetical dilemmas.

Interestingly, my results showed greater overall cooperation in the gains context relative to the losses. This is consistent with other work supporting a self-oriented perspective on loss aversion in social dilemmas, whereby an aversion to losses makes people demand more in interpersonal interactions (De Dreu et al., 1994, 1992), increases cheating (Schindler & Pfattheicher, 2017), and reduces prosocial behaviour (Brewer & Kramer, 1986). On the other hand, this is perplexing because it contrasts both with more recent work taking an other-oriented perspective on loss aversion (e.g. Leliveld et al., 2009) and the findings reported in the previous chapter whereby there appeared to be higher cooperation in the losses frame than the gains frame. Why might this be so?

One possibility for why I observed that the loss frame decreased prosocial behaviour is that my loss-framed PD in this chapter was more cognitively complex than the loss-framed ILT in Chapter 2, both because of the interdependent structure and the mental calculations involved. With regards to interdependence of outcomes, unlike the

ILT the PD is an interdependent game. In the PD, participants must consider not just what they themselves would like to do, but also what the other person might choose and how that dynamic would influence their own choice. In the previous chapter, participants only had to consider what they themselves were motivated to do, in the absence of any fear motive. Similarly, in the losses frame here participants must remember that they had a starting payment of $0.80 and then mentally subtract the values in the matrix from that. In contrast, the payoff structure and exact payments for each outcome in each trial were displayed in Chapter 2, reducing the need for mental calculations. Given this, especially if more deliberative, strategic thinking is associated with self-interested behaviour, this could explain why there was less cooperation in the losses frame in this study. On the other hand, if the losses frame was more cognitively complex one might expect there have been an interaction between gains/loss frame and the time constraint manipulation – which there wasn't. Another, not mutually exclusive, possibility is that because this study used an interdependent PD, participants might have expected, in a kind of second-order loss aversion, that other people would be more likely to be selfish and so participants coordinated their own outcomes accordingly. Because I did not measure explicit expectations of behaviour, it is impossible to know from the current data and so this remains a point for future research.

Using Time Constraint Studies to Investigate Intuitive Cooperation

There has been a growing awareness of methodological problems with the use of time pressure manipulations in studying intuitive prosociality (e.g. Bouwmeester et al., 2017; Tinghög et al., 2013), and these have led some researchers to suggest that the SSH

is based on fundamentally flawed evidence. Intuitive cooperation, it is argued, is a mere experimental artefact. My third contribution in this chapter is to develop a new design that addresses these methodological concerns while still observing the basic effect of the SSH.

First, I drastically reduced participant exclusions from around 50% to just 2%, and made them evenly distributed across the two manipulation conditions. Second, I ensured that all participants understood the structure of the game. That even when addressing these problems I observed greater cooperation in the time pressure condition relative to the time delay condition helps show that intuitive prosociality cannot merely be seen as an experimenter artefact, thus giving confidence to researchers working on intuitive cooperation and ameliorate concern about selection bias in existing studies. That said, this does not mean that time pressure is the only – or even the best – experimental method to manipulate reliance on intuition vs. deliberation. While time pressure is a good measure in that it is straightforward and easily implemented even in online studies, Rand's (2016) meta-analysis suggests that the effects of intuition on cooperation were weaker when implemented through time pressure than when using alternative methods such as conceptual priming or cognitive load. Unfortunately, however, there were not enough studies using these alternative methods to compare between them. It would be useful for future work to look in more detail at the efficacy of studying intuition through time pressure or alternative methods.

The third methodological improvement I made was the inclusion of an unconstrained control condition. Previous work has typically ignored what an

individual's 'default' response mode is when not instructed either to rely on intuition or to second-guess their first guess and deliberate. By including a neutral control condition, I could look at how participants behaved in the absence of time constraint instructions, thus providing further nuance by demonstrating that intuitive cooperation effects seem driven more by time-delay induced deliberation reducing cooperation than by time-pressure induced intuition increasing cooperation (though neither were significantly different from the control condition at the $p < .05$ level). This is entirely consistent with the SSH, which varying describes the effects of such time constraint studies as both indicating "intuitive cooperation" and "self-interested deliberation". But what my results do imply is that more of the action might be coming more from self-interested deliberation than intuitive cooperation. While this does not matter at a practical level when intuition and deliberation are seen as opposite ends of a single continuum, it does suggest that work on the application of dual process models should pay greater attention to this potential disparity. People's baseline is to cooperate, but that this can be increased or decreased depending on possibility of deliberation.

Limitations

There are a few more limitations of the present work that deserve mention. First, while I do have a large sample size with around 100 participants per cell and did conduct an a priori power analysis to ensure sufficient power to detect moderate effects, this still might not have been adequate to detect small two-way or three-way interactions. Testing this in future might require the use of mixed designs or even larger samples.

Second, we used deception in this task – both in the group manipulation and by not paying participants according to their decisions. This is one feature that differentiates my work from previous studies (e.g. Bouwmeester et al., 2017) and it would be preferable in future work for a full non-deception procedure to be implemented. I return to this issue of deception in my general discussion in the last chapter.

Third, a potentially important limitation is that I used an MTurk sample. While evidence suggests that MTurk does yield reliable samples that are more representative than college students (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Hauser & Schwarz, 2016), there are certain issues that occur when using such online samples. Like for deception, I shall discuss this issue of using MTurk samples at more length in my final chapter.

Fourth and finally, I used an arbitrary, competitive groups paradigm and it remains possible that a more forceful manipulation drawn from real social groups would yield a different result (e.g. an interaction between time constraint and group membership). I note, however, that my group manipulation was strong enough to yield a significant main effect of group membership, such that people cooperated more with members of their own group across all conditions.

Conclusion

In conclusion, in this chapter I studied the effects of intuitive ingroup cooperation, developing a new procedure that addresses methodological concerns that have plagued previous work. My results show that ingroup cooperation and ingroup favouritism emerge even when not given the chance to deliberate, supporting the idea that intergroup

bias runs deep in human cognition. We might be intuitive cooperators, but not intuitive egalitarians.

Chapter 4: Reciprocity and Religious Prosociality

This chapter is an extended version of a paper published as:

Everett, J. A. C., Haque, O. S., & Rand, D. G. (2016). How good is the Samaritan, and why? An experimental investigation of the extent and nature of religious prosociality using economic games. *Social Psychological and Personality Science*, *7*(3), 248–255. https://doi.org/10.1177/1948550616632577

Abstract

How does ingroup-favouring prosocial behaviour in real groups manifest in contexts that either contain or lack interdependence of outcomes and expectations of reciprocity? Looking at groups based on religious belief, I examined how much of a $0.30 bonus MTurk workers would share with the other player whose religion was prominently displayed during two online resource allocation games. In one game (but not the other), the recipient could choose to reciprocate. Results from both games showed that more religious individuals were more prosocial in both games towards both atheists and Christians. While atheists showed ingroup favouritism, Christians did not.

Introduction

As Allport (1954) noted in his highly influential *The Nature of Prejudice*, the role of religion is paradoxical: "It makes prejudice and it unmakes prejudice" (p.444). The same could be said for religion's effect on prosocial behaviour. On the one hand, history books are full of stories of those whose altruism was inextricably linked to their religious conviction: Francis of Assisi; Martin Luther King Jnr; the Dalai Lama. For much of history the poor were provided for by religious organisations, and many of the movements for social progress – such as William Wilberforce's abolitionist movement - were driven by religious conviction. These are, however, not the only stories that one finds in the history books. The books are also full of those who have justified tremendous suffering in the name of religion. One finds, for example, Pope Urban II, who issued the decree for the First Crusade in which the leaders of Christendom (as Europe was then known)

temporarily stopped killing each other and instead marched upon Muslims living in the 'Holy Land'. In modern times, one is regularly confronted by violence inflicted in the name of God by the Islamic State, or the hate-filled protests and picketing of funerals by the Westboro Baptist Church. Such connections have led some to conclude, not entirely unreasonably, that religion is "violent, irrational, intolerant, allied to racism and tribalism and bigotry" (Hitchens, 2007, p. 56).

When, for my third chapter, I wanted to look at prosocial behaviour in 'real' rather than arbitrary experimenter-induced groups, religious groups were a natural choice. Christians and atheists in the U.S. are often in conflict (Edgell, Gerteis, & Hartmann, 2006), and there is a tension that exists between the pacifistic, universal aspects of the Christian faith and the more parochial ones. Studying how religious individuals make prosocial decisions in intergroup contexts helps elucidate not only the factors that can influence prosocial behaviour in general, but also those that can influence parochial prosociality specifically.

Expectations of Reciprocity in Ingroup Favouritism

To what extent is ingroup favouritism observed even in non-interdependent contexts where there can be no expectation of reciprocity? In Chapter 2 I presented evidence that even in a context where there is no interdependence of outcomes, no direct expectations of reciprocity, and no potential for reputation-building, minimal group members show ingroup favouritism. This is consistent with ingroup favouritism being driven, at least in part, by 'mere categorization' and social identity processes. However, my work in Chapter 2 involved minimal groups, which differ in important ways from

real groups, and I did not include a game in which there was any direct interdependence of outcomes and the direct potential for reciprocity. Indeed, minimal groups do differ (intentionally) from real groups in a number of ways, and the original intention of using minimal groups was to look at how the simple act of categorization could produce intergroup behaviour, even when there is no "conflict of interest nor previously existing hostility between the groups" and no "rational link between economic self-interest and the strategy of ingroup favouritism" (Tajfel & Turner, 1979, p. 38). What about when there is pre-existing hostility between groups, and when there is interdependence of outcomes? For this reason, in Chapter 4 I moved away from the gains and losses paradigm to look at real groups that have a real history of intergroup conflict and tested behaviour both in a Dictator Game (DG: Forsythe et al., 1994) and Trust Game (TG: Berg et al., 1995). Comparing these two games helps to highlight the difference between a more strategic, self-interested, and expectation-based route to ingroup favouritism (as in the TG) and a more non-strategic, other-regarding route to ingroup favouritism (as in the DG).

A TG typically has two participants: an investor and a trustee. The investor is given some money and told that they must send a proportion (from zero to the full amount) of this money to the trustee, and that the experimenter will multiply the money by some amount. Once the trustee receives the money, they are told that they must send back a portion of it to the investor, again ranging from zero to the full amount. Thus, if the investor has an expectation of reciprocity and trusts that the recipient will return 50% or more of what they receive, the investor pays no cost (or even makes money) while helping the recipient. Critically, then, outcomes in the TG are interdependent (like the PD in

Chapter 3) because the final payoff of the investor depends upon the behaviour of the trustee. It is this interdependence of outcomes that opens the door for prosocial behaviour driven by relatively more strategic and self-interested expectations of direct reciprocity.

The TG can be compared with the simpler variant that is the DG. In the DG one player - the dictator – is given some money and makes a unilateral decision about how much of this money to share with a second (anonymous) player - the recipient - who must accept whatever amount the dictator chooses to give. Behaviour in an anonymous one-shot DG can be interpreted as resulting from relatively more other-regarding motives, rather than being self-interested (because acting prosocially will always lead to a worse monetary outcome for the self) or driven by expectations about how the recipient will respond (because the recipient does not take any action there is no potential for reciprocity). Of course, given 'spill-over' effects whereby people adopt strategies used in one game to another, behaviour in the DG likely involves some residual reciprocity-based processes even though there is no reciprocity available in this specific game. Nonetheless, given the inherent structure of the games one can be reasonably confident that in the DG, expectations of reciprocity are *less* important than in the TG, even if they are not eliminated. This contrast between the games sets up an ideal way with which to look at expectations of reciprocity in intergroup prosocial behaviour. If expectations of reciprocity are essential for ingroup favouritism to occur, one should expect ingroup-favouring prosocial behaviour to occur only in the TG and not an anonymous DG. In contrast, if expectations of reciprocity are an important, but not necessary, condition for ingroup favouritism to occur one would expect ingroup-favouring prosocial behaviour

to occur both in the TG and the DG. The groups that I choose to use to explore this in were Christians and atheists.

Religious Prosociality

In popular perception, belief in God is inextricably linked with being good. Religious individuals are rated as being more moral than their nonreligious counterparts (Hout & Fischer, 2002) and working professionals who show religious beliefs are rated as more likable, trustworthy, and moral than their nonreligious counterparts (Bailey & Doriot, 1985). In contrast, those who do not believe in God are typically perceived as "self-interested individualists who are not concerned with the common good" (Edgell et al., 2006, p. 225), and consequently atheists experience distrust and prejudice - particularly from religious people (Gervais, Shariff, & Norenzayan, 2011). The idea that religiosity is associated with prosociality remains powerful in society.

Is there truth in the stereotype? Throughout history, giving to charity has been required by some religions – for example, *tithing* in Christianity, and *zakah* in Islam. Moreover, contemporary sociological surveys in the US have supported the idea that religious people act more prosocially. Religious states in the US give more to charity than nonreligious states (Brooks, 2007), and even when controlling for other factors such as education, gender, income, and politics, those who frequently pray and attend religious services reliably report more prosocial behaviour such as charitable donations and volunteerism (Brooks, 2007; Monsma, 2007; Putnam & Campbell, 2012). There are, however, important problems with the use of self-report measures to study prosociality in general, and religious prosociality specifically. Individuals tend to over-report

prosocial actions based on introspective intentions (Batson, 1991) and it is possible that religious people are more likely to over-report compared to the nonreligious because of an internalization of the stereotype that religious people actually are more prosocial (Galen, 2012). Consistent with this, the relationship between religiosity and prosociality does appear substantially stronger when looking at self-report measures than it does when looking at behaviour measures (Galen, 2012). It is critical, therefore, that when investigating the relationship between religion and prosocial behaviour, one looks at actual behaviour. An ideal way to measure such prosociality comes from the use of behavioural economic games wherein participants make real monetary choices that affect earnings for themselves and others. Unfortunately, existing evidence using such measures is yields mixed results.

Generalized, Parochial, or Non-Existent?

In line with what one might expect given the importance of charity and prosocial behaviour in religious systems, some studies report a positive relationship whereby religiosity is associated with prosocial behaviour as measured by giving in DGs (Eckel & Grossman, 2003; Tan, 2006), cooperation in PGDs (Ahmed, 2009), and trustworthiness in TGs (Tan & Vogel, 2008). For example, Ahmed (2009) has presented evidence from a rural Indian community showing that religious students were more cooperative and prosocial in both PGD and DGs. Supporting the idea that religious people are perceived as more trustworthy, Tan and Vogel (2008) show that in a TG religious trustees are trusted more (particularly so by other religious participants), and were in fact more trustworthy than atheists (Tan & Vogel, 2008).

However, other researchers have found - more consistent with the idea that religion is not a force for good – no relationship between religious belief and prosocial behaviour. For example, Orbell et al. (1992) find that although both religious and non-religious people believe that the former are more cooperative than the latter, there is – in actuality - no relationship of religiosity and giving in a PGD. Other scholars have suggested that there is no relationship of religiosity with giving in DGs (Eckel & Grossman, 2004; Paciotti et al., 2011), trustworthiness in TGs (Anderson, Mellor, & Milyo, 2010; Bellemare & Kröger, 2007), or cooperation in PGDs (Ahmed & Salas, 2009; Anderson & Mellor, 2009; Anderson et al., 2010; Orbell, Goldman, Mulford, & Dawes, 1992).

Perhaps most intriguing from the perspective of my thesis is the link between religious belief and intergroup bias. Ingroup favouritism is a robust phenomenon, observed for a whole range of groups and in many different situations and there is some evidence suggesting that religious people might even be especially likely to exhibit intergroup bias. For example, research has documented small but significant correlations between at least some forms of religiosity and some forms of prejudice, including racism (Hall, Matz, & Wood, 2010) and homophobia (Herek, 2000). When it comes to prosocial behaviour specifically, some research suggests that religious prosociality is deeply parochial. For example, it has been found that Kibbutz members in Israel show higher contributions in a PGD when paired with other Kibbutz members than with anonymous city members (Ruffle & Sosis, 2006). Similarly, Christians in New Zealand are more

generous to anonymous Canadian Christians than New Zealand citizens are to other anonymous New Zealand citizens (Bulbulia & Mahoney, 2008).

Consideration of the existing literature on religious prosociality, then, reveals mixed effects such that three competing hypotheses can be discerned. First, a religious *prosociality equivalence* hypothesis predicts that religious and non-religious individuals are equally prosocial. Second, a *generalized religious prosociality* hypothesis predicts that religious individuals are more prosocial regardless of the recipient. Third, a *religious parochialism* hypothesis predicts that religious people are only more prosocial towards co-religionists.

In this chapter I wanted to distinguish between these competing hypotheses by examining the relationship between religiosity and prosociality when giving to Christian versus atheist partners in economic games. Previous work has often compared prosociality among religious individuals vs. individuals that have not specified a religion (e.g. Kibbutz members vs. anonymous city members: Ruffle & Sosis, 2006), but this is problematic when one considers that in the U.S. around 89% of people believe in God, and so when not given evidence to the contrary individuals are likely to assume someone does believe in God (Galen, 2012). In other words, religious prosociality could still be parochial even in the studies from which it is presumed to be generalized because participants are perceiving unidentified individuals as religious. For a comparison group to study the extent of parochial religious prosociality, atheists are an ideal recipient group: they constitute a clearly distinct group from religious individuals (one cannot easily be a Christian *and* an atheist), and one that is particularly disliked across many

demographics (Edgell et al., 2006). If religious prosociality is primarily parochial – rather than generalized – one would expect prosocial behaviour to be extended selectively towards other religious individuals and not atheists. On the other hand, if religious prosociality is a generalized phenomenon, one would expect religious individuals to be more prosocial even to members of such a generally disliked group.

Why would religious people be more prosocial?

Leaving aside disagreements in the literature as to *whether* and to *whom* religious people are more prosocial, different accounts have also been posited to explain *why* religion might facilitate prosociality. One possibility is that religious individuals may be more prosocial due to the kinds of beliefs they have about the nature of reality, whereby a sensitivity to punishments and rewards from supernatural agents may enforce norms even in the absence of human social monitoring (Atkinson & Bourrat, 2011; McKay, Efferson, Whitehouse, & Fehr, 2011; Norenzayan & Shariff, 2008). Alternatively, religious prosociality may be a by-product of sociological factors, including stronger social networks among religious groups (Lewis, MacGregor, & Putnam, 2013; Putnam & Campbell, 2012). Yet more possibilities are that specific kinds of religious practices inculcate prosocial tendencies, such as religious rituals (Xygalatas, 2013) or religious experiences (James, 1902/1985). To help shed light on this more exploratory question, I measured multiple components of religiosity (including frequency of private prayer, frequency of public worship, and certainty of belief in God) to see, if religious prosociality is observed, what elements of religiosity best predict this prosociality.

In the present research, then, I studied ingroup favouritism in real groups by looking at behaviour in both a TG – a game where transfers are driven by expectations of reciprocity - and a DG – a game where there is no potential for reciprocity and so decisions result more from relatively non-strategic prosocial motives. By using religious groups to study this question, I could use this opportunity to gain a deeper insight into the nature of religious prosociality and help resolve lingering questions as to whether religious prosociality is generalized, parochial, or even non-existent.

Study 1

Method

Participants

582 American participants (268 female) were recruited online using Amazon Mechanical Turk and paid $0.50 for their time, with an option of keeping an additional bonus of $0.30. Only participants who passed simple comprehension checks concerning the payment structure and rules of the games were included in data analysis[38]. Because I had planned on running a TG in a subsequent study and the TG instructions are more complex than those of the DG, I was concerned that there would be systematic variation in participants across studies (i.e. participants in the TG would be more sophisticated on

[38] While this chapter and the following one are presented in this thesis after the work reported in the previous chapter in the interest of presenting the work in a more linear fashion, this was not the order they were conducted chronologically. For this reason, the enhanced attention and comprehension measures developed, discussed, and used in Chapter 3 were unfortunately not used in this chapter.

average than those in the DG). Therefore, after participants made their DG decision, they were shown the instructions for the TG and completed the TG comprehension checks.

In this chapter, results are reported only for participants who passed comprehension checks for both games, took the survey only once, and who identified as either Christian or Atheist (N = 296). Supplementary results including all religious participants can be seen in Appendix A, and results did not substantially change based on these exclusion criteria. Of the Christian participants (N = 165), nearly all were at least moderately certain that God exists (98%); and with a frequency of at least "occasionally": 83% reported thinking about religious issues; 80% reported praying; 76% reported experiencing God intervening in their life; and 55% took part in religious services. Of the Christian participants, the most common denomination reported was the Catholic Church (30%), followed by the Southern Baptist Convention (15%), United Methodist (7%) and Presbyterian (7%). Participants were politically moderate on average (on a 1-7 scale; M = 4.48, SD = 1.67).

An *a priori* power analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) revealed that for the a 2x2 design and taking Cohen's f effect size standards of 0.14 as small, and 0.39 as medium, 256 participants were required to detect a small-to-medium sized interaction effect (f =0.20) with an alpha of .05 and power of .80. Therefore, this study was sufficiently powered.

Design

This study had a 2 (Participant Religiosity: Christian vs. Atheist) x 2 (Recipient Religiosity: Christian vs. Atheist) between-subjects design, such that Christian or atheist participants played a DG with a partner who was either Christian or atheist.

Participants were informed at the start of the study that they would be playing a simple game with another participant. To manipulate the religious identity of the recipient (Person B; the second player), participants were given four pieces of information about the other player. To eliminate reputational concerns, participants were explicitly reminded that the other player would not know any information about themselves or their (lack of) religious affiliation. The other player was described as living in the United States, speaking English as their primary language, and being over the age of 18. The fourth piece of information constituted the experimental manipulation, where participants were told that the player identified as Christian or as an atheist. This information was presented in both text form and with a symbol presented on both the pre-decision page and the decision page. It was made salient to participants that there was no deception in this task and that the decisions they made were real. Participants answered questions on their own religiosity after making their decisions in the economic game to avoid potential religious priming effects. Participants were given the following instructions:

> "In this interaction you are matched with a real other person. One of you will be Person A, one of you will be Person B. Person A starts with a 30 cent bonus, and Person B starts a 0 cent bonus. Person A makes a choice. Person

> A can choose how much of their 30 cents to transfer to Person B. Whatever person A transfers is doubled and given to Person B."

Participants indicated how many cents they wanted to transfer in 5-cent increments. I doubled the money transferred in my DG in order to make the game equivalent to the Trust Game, described below in Study 2, with the important exception that in the DG the recipient is powerless and takes no action, therefore precluding any positive return of the doubled money. As I shall return to in the limitations section of this chapter, however, this methodological change could have been a problem by making the prosocial behaviour more efficient than it would be in a typical DG.

Measures

Centrality of Religion was measured using Huber and Huber's (2012) short item Centrality of Religiosity Scale (CRS). The CRS is a measure of the centrality, importance, and salience of religiousness in a person, and consists of five items that each tap one of the theoretically defined core dimensions of religiosity: public practice, private practice, religious experience, ideology and intellectual. Participants were asked to rate on a 5-item scale "How often do you think about religious issues?" (Intellect: *1 = never; 5 = very often*); "To what extent do you believe that God or something divine exists?" (Ideology: *1 = not at all, 5 = very much so*); "How often do you take part in religious services?" (Public practice: *1 = never, 5 = very often*); "How often do you pray?" (Private practice: *1 = never, 5 = very often*); and "How often do you experience situations in which you have the feeling that God or something divine intervenes in your life?" (Experience: *1 = never, 5 = very*

often). Scores were combined into a single reliable measure (α = .91) of the centrality of religiousness for each participant.

Negative Attitudes Towards Atheists were measured using the Negative Attitudes Towards Atheists (NATA) scale (Gervais, 2011). Participants rated their agreement with 7 statements on a 1-7 scale (*1 = not at all, 7 = very much*), including "I would prefer to spend time with people who are religious believers" and "I would be uncomfortable with an atheist teaching my child" (α = .92). Given that I use an American sample here rather than the Canadian sample with which it was originally used, I changed the item "I would not at all be bothered by a Prime Minister who did not have religious beliefs" to instead refer to a President.

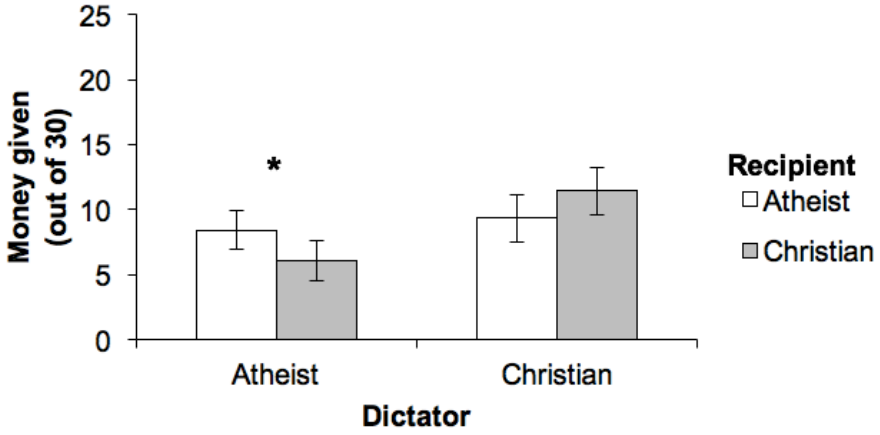
Analysis

For data analysis, I used a regression-based analysis procedure to examine the relationship between the continuous (Participant religiosity on the CRS) and categorical (Recipient religiosity: Atheist vs. Christian) predictors and the dependent measure of behaviour in the economic game. For each analysis, I first looked at the main effects of the centered continuous variable (participant religiosity) and the categorical manipulated variable of recipient religiosity (Atheist recipient coded as -1, and religious recipient coded as +1). For the interaction, I added the interaction term of the categorical and continuous variables and looked at simple effects by evaluating the effect of recipient religiosity at +/- 1 SD of participant CRS.

Results

I first looked at whether there were main effects of participant and recipient identification and an interaction effect between the two, indicating ingroup favouritism. Looking at the binary identification measure of participant religious belief, Christian participants transferred more in the DG, $F(1,292) = 12.39$, $p < .001$, and there was a significant interaction, $F(1,292) = 6.28$, $p = .01$, such that only atheist participants, $t(128.41) = 2.19$, $p = .03$, but not Christian participants $t(163) = -1.58$, $p = .12$, transferred significantly different amounts based on the recipient's religiosity (see **Figure 4.1**). That atheist participants showed ingroup favouritism in this anonymous DG is consistent with my results from Chapter 2 where ingroup favouritism can be observed even in contexts where there can be no potential for reciprocity or reputation building. Interestingly, however, religious participants did not show ingroup favouritism and were more prosocial regardless of the target.

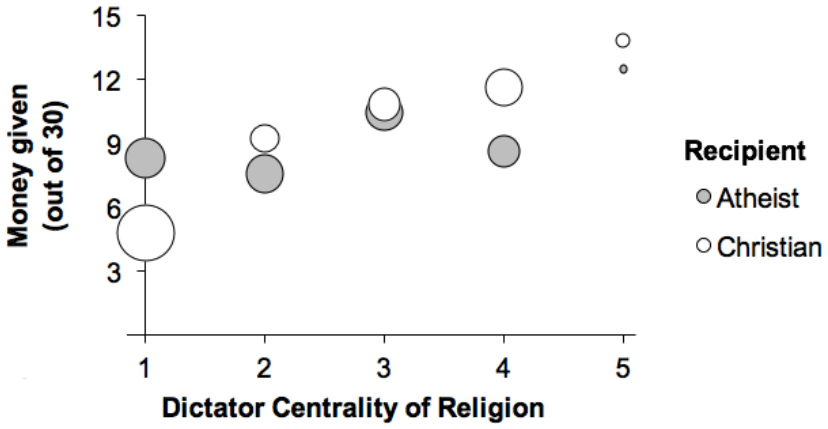
Figure 4.1: Transfers in cents as a function of self-identification as Christian or Atheist and target religiosity. Error bars represent 95% C.I.s



I complemented these analyses with a regression-based approach to explore further this finding from the binary identification analysis that religious people were more prosocial regardless of the recipient. To do this, I looked at the continuous measure of participant religiosity (CRS from 1 to 5, centered), as well as recipient religiosity (categorical manipulated variable; atheist (-1) or Christian (+1)), and the interaction between the two. As before, there was a significant main effect of participant religiosity, $B = 1.73$, $t(293) = 4.94$, $p < .001$, whereby religiosity was associated with increased transfer amounts (see **Figure 4.2**) and no significant main effect of recipient's religiosity, $B = 0.02$, $t(293) = 0.05$, $p = .96$. Perhaps unsurprisingly given that this analysis does not include *identification* as atheist or religious but only degree of religiosity, the interaction between continuous participant religiosity and recipient religiosity did not reach statistical

significance[39], B=0.67, $t(292) = 1.89$, $p = .06$. These effects were robust to controlling for gender, ethnicity, education, and political ideology; held when looking only at Christians and not atheists; and held when using participants from all religious affiliations (see Appendix A for statistics). In all three cases, the marginally significant interaction between participant religiosity and recipient religiosity became fully non-significant. Individuals with stronger religious beliefs were significantly more prosocial to both Christian and atheist recipients; the evidence does not support a religious parochialism hypothesis.

Figure 4.2: Transfers in cents as a function of recipient and participant religiosity (binned to the nearest integer with the mean displayed across all observations in that bin). The size of the dots is proportional to the number of observations.



[39] When I looked at the simple effects (given that $p = .06$ and the binary identification showed an interaction), I found no significant effect of recipient religiosity for either more or less religious participants.

Further considering my finding that religious people were more prosocial, I next conducted these analyses using each individual component of the CRS to test whether these effects were explained by one particular component of religiosity. While there were significant main effects of all five components of the CRS when analysed separately, follow-up mediation analyses indicated that how often participants thought about religious issues fully mediated the relationship between DG transfers and all four other components of the CRS (extent of belief in God, frequency of religious practices or experiences). Furthermore, a multiple regression analysis including each CRS component together showed that the frequency with which participants reported thinking about religious issues was the only significant predictor, $B = 1.78$, $t(289) = 3.69$, $p < .001$ (see Appendix A). Thus, the extent to which participants believed that God exists, and the frequency of religious experiences, prayer and attending religious services did not appear to influence their level of prosociality, except in so much as they acted as proxies for frequency of thinking about religious issues. Furthermore, I note that when replicating my main analysis using frequency of thinking about religious issues rather than the full CRS measure, there is even less evidence of an interaction between participant religiosity and recipient religiosity, $B = 0.47$, $t(293) = 1.25$, $p = .21$.

Finally, to ensure that these results were not driven by the sample being unusually unprejudiced towards atheists, I explored whether self-reported negative attitudes towards atheists translated into actual discrimination. I found a significant positive main effect of NATA on giving, $B = 0.92$, $t(293) = 3.41$, $p < .001$, but no significant main effect of recipient religiosity, $B = 0.04$, $t(293) = 0.09$, $p = .93$, and no interaction between the two,

B = 0.32, $t(292)$ = 1.17, p = .24. Thus, surprisingly, self-reported anti-atheist prejudice actually led to *increased* giving to atheists, and did so just as much as to religious recipients. This relationship between NATA and transfer amounts was fully mediated by participant religiosity, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 1.33) not including zero (0.50 – 2.17). Thus, it was increased religiosity rather than increased anti-atheist prejudice which led to increased giving to both atheist and religious recipients.

Study 2

Method

600 participants (282 female) were again recruited through MTurk in the same between-subjects design as in Study 1. The instructions and response format for the TG were designed to be as similar to the DG as possible, where participants chose how much of an initial $0.30 endowment they would like to transfer to a second participant who was either atheist or religious. The crucial difference was that in the TG participants were told "Person B can then choose how much of the money they want to transfer back to Person A". Participants indicated how much they would like to transfer in 5-cent increments, and after the transfer were asked to indicate on the next page the percentage of the money they expected the second player to transfer back to them (0%, 25%, 50%, 75%, 100%). Predicted returns were measured to complement the main measure of transfer amounts in the TG as a more explicit measure of trust in the recipient.

In order to eliminate deception[40] and determine the bonuses paid to participants, I looked at the decisions of two individuals who played as the second mover in a TG in an earlier unrelated study; identified either as a Christian or Atheist; and matched the demographic information told to participants. These two second movers were then matched with all participants from the current study, and paid accordingly for each matching. As the instructions did not indicate that second movers played only one game, this multiple matching procedure was not inconsistent with the instructions and therefore did not constitute deception.

As in Study 1, results are reported only for participants who passed comprehension checks for *both* games, took the survey only once, and who identified as either Christian or Atheist (N = 272), but results using all religious participants can be seen in the Appendix. Of the Christian participants (N = 140), nearly all were at least moderately certain that God exists (97%); and with a frequency of at least "occasionally", 83% reported thinking about religious issues, 79% reported praying, 79% reported experiencing God intervening in their life, and 62% took part in religious services. Participants were, again, politically moderate on average (M = 4.67, SD = 1.72). An *a priori* power analysis using G*Power (Faul et al. 2009) revealed that for the a 2x2 design and taking Cohen's f effect size standards of 0.14 as small, and 0.39 as medium, 256

[40] This no-deception practice differs from the previous two studies, in large part due to the differences in methodological practice at different institutions and in different research groups within those institutions. This work was conducted with David Rand, who adopts a no-deception policy across all studies. In social psychology at Oxford, where I ran the experiments in Chapter 2, deception is much more common. I shall return to this issue of deception in my last chapter.

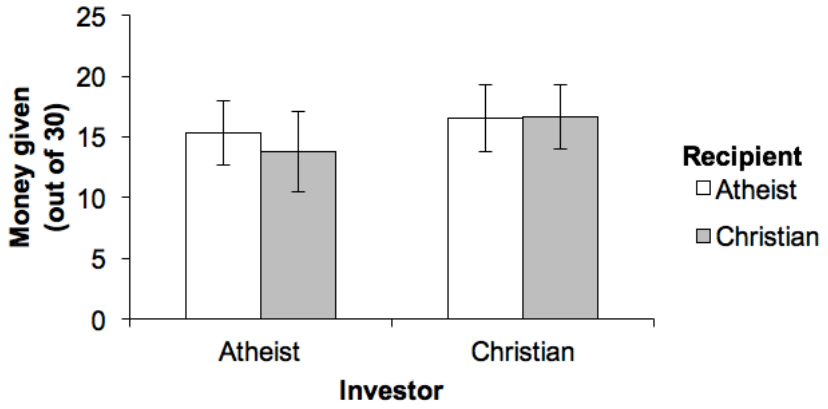
participants were required to detect a small-to-medium sized interaction effect (f =0.20) with an alpha of .05 and power of .80.

Results

Transfer Amounts

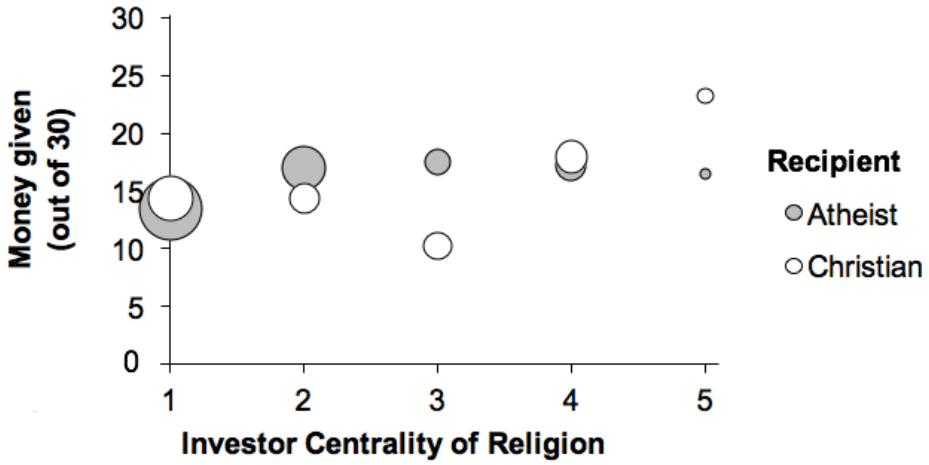
Using the binary identification measure, and contrary to the results in the DG, Christian participants did not transfer more in the TG, $F(1,268) = 2.01$, $p = .16$. There was no effect of recipient's religiosity $F(1,268) = 0.25$, $p = .62$, and there was no interaction with recipient religious belief, $F(1,268) = 0.35$, $p = .56$ (see **Figure 4.3**). In other words, Christians or atheists did not show ingroup favouritism in the TG and participants who identified as Christians did not transfer more on average than participants who identified as atheists. When using a regression-based approach with the continuous measure of religiosity, there was a significant main effect of participant religiosity, $B = 1.46$, $t(269) = 2.62$, $p = .009$, whereby religiosity was associated with increased transfer amounts (see **Figure 4.4**). As with the binary measure, there was no significant main effect of recipient's religiosity, $B = -0.43$, $t(269) = -0.61$, $p = .55$ or interaction between the two, $B = 0.39$, $t(268) = 0.70$, $p = .48$. Religious participants, in order words, transferred more in the TG. This effect using the continuous measure of religiosity was robust to controlling for gender, ethnicity, education, and political ideology; looking only at differences within Christians and not atheists; and held when using participants from all religious affiliations (see Appendix for statistics). It remains to be seen why identification as a Christian was not a significant predictor of transfer amounts, but crucially, the continuous measure was – including when looking at religious differences only within self-identifying Christians.

Figure 4.3: Transfers in cents as a function of self-identification as Christian or Atheist and target religiosity. Error bars represent 95% C.I.s



Next, I conducted these analyses using each individual component of the CRS to test whether these effects were explained by one particular component of religiosity. As in Study 1, there were again significant main effects of all five components when analysed separately. However, frequency of thinking about religious issues fully mediated the effect of all other CRS components, and was the only a significant term in a multiple regression including each of the five components, B = 1.91, $t(265)$ = 2.45, p = .01. In contrast to Study 1, I found no main effect of NATA on giving, B = 0.72, $t(269)$ = 1.62, p = .11, no significant main effect of recipient religiosity, B = -0.45, $t(269)$ = -0.63, p = .53, and no interaction between the two, B = 0.32, $t(268)$ = 0.70, p = .48.

Figure 4.4: Transfers in cents as a function of recipient and participant religiosity (binned to the nearest integer with the mean displayed across all observations in that bin). The size of the dots is proportional to the number of observations.



Predicted Returns

I next used the same analyses to explore the effects of participant and recipient religiosity on participants' expectations regarding how much their partner would return to them in the TG. Looking at the binary identification measure, there was a main effect of participant identification such that Christian participants expected more to be returned in general than atheist participants did, $F(1,268) = 4.65$, $p = .03$, but there were no main effects of recipient religiosity, $F(1,268) = 0.54$, $p = .46$, and no interaction effect, $F(1,268) = 1.43$, $p = .23$. Looking at the continuous measure of religiosity, the same main effect of participant religiosity was observed such that more religious people expected more to be returned, $B = 3.52$, $t(269) = 3.03$, $p = .003$. There was no main effect of recipient religiosity, $B = -1.19$, $t(269) = -0.79$, $p = .43$, but interestingly there was this time an interaction effect, $B = 2.27$, $t(268) = 1.96$, $p = .05$, such that religious participants showed no difference in

expectations based on the partner religious belief, $B = 1.70$, $p = .42$, but atheist participants trusted religious recipients to transfer significant less than atheist ones $B = -4.12$, $p = .05$. That is, overall religious participants predicted more to be returned, which is what would be expected given the structure of the game and the previous finding that religiosity was associated with transferring more. It is interesting that despite not transferring less, participants that were less religious were more distrustful of religious recipients specifically.

Discussion

In this chapter I studied ingroup-favouring prosocial behaviour in real groups by looking at behaviour of Christians and atheists in two games that either had or lacked interdependence of outcomes and the potential for reciprocity. By using religious groups to study this question, I could also use this opportunity to gain a deeper insight into the nature of religious prosociality and help resolve whether religious people are only parochially prosocial towards co-religionists (e.g. Ruffle & Sosis, 2006), more prosocial in general (e.g. Ahmed, 2009) or not even more prosocial at all (e.g. Eckel & Grossman, 2004). I looked at how much Christian and atheist participants transferred to Christian and atheist recipients in an anonymous DG (a game with no reciprocity and no potential for reputation-building) and TG (a game in which interdependence and expectations of reciprocity are important). My results showed that atheist-identifying participants (but not Christians) showed ingroup favouritism and transferred more to fellow atheists in the DG (but not TG). In contrast, Christian participants transferred more money to both Christian and atheist recipients in both a DG and TG.

Contributions

A first contribution of my work here is to corroborate the findings of Chapter 2 in demonstrating that ingroup favouritism can be observed even in anonymous contexts without the potential for direct reciprocity or reputation building. I found that self-identified atheist participants showed ingroup favouritism by giving higher amounts in an anonymous DG to an atheist recipient than a Christian one (even though this was still less than the amount Christians transferred to atheists). As I argued in my second chapter, that (atheist) participants showed ingroup favouritism even in an anonymous game where there is no chance of reciprocation or reputation-enhancement is inconsistent with the BGR model and instead highlights the importance of social identity. Interestingly, there was no such ingroup favouritism in the TG in Study 2. This suggests that while atheist participants may have had stronger preferences to help an ingroup member, they didn't expect that fellow atheists would be more trustworthy (and, again, overall atheist participants were less trusting than Christian participants regardless of the other player's belief).

Second, and somewhat in contrast, I found no ingroup favouritism in Christian participants: they were more prosocial in both the DG and TG to both Christian and atheist recipients. This increased religious prosociality was explained entirely by how often people thought about religious issues. Such results provide clear behavioural evidence supporting conclusions from correlational and self-reported sociological data on the positive relationship between religiosity and both formal and informal prosocial behaviours (e.g. Monsma, 2007; Putnam & Campbell, 2012). At the same time, my results

also suggest that religious prosociality effects cannot be explained entirely as a result of social networks (e.g. Lewis et al., 2013; Putnam & Campbell, 2012), accountability arising from belief in an all-knowing God who punishes bad actions and rewards good actions (e.g. Atkinson & Bourrat, 2011; McKay et al., 2011), or the specific nature of religious rituals or experiences (e.g. James, 1902/1985; Xygalatas, 2013). How, then, might this lead to reduced ingroup favouritism? This is not entirely clear, and is in fact potentially quite surprising given that ingroup favouritism is typically observed in a variety of contexts. One possibility is that this reflection on religious issues represents reflection on moral norms within religious traditions that decree how to live. It is possible that this reflection helps inculcate internalized social heuristics, in a similar way to how behavioural experiences shape tendencies to act in prosocial ways, as suggested by both interdependence theory (e.g. Kelley et al., 2003; Van Lange et al., 1997) and the SSH (Rand, 2016; Rand et al., 2014). Indeed, that the experiences of religious individuals might lead to an internalization of social heuristics is consistent with findings in some of my own work not included in this thesis, in which I have found evidence that social gender roles differentially influence the intuitive behaviour of men and women in the DG because of the prior experiences men and women (Rand, Brescoll, Everett, Capraro, & Barcelo, 2016). It will be interesting to explore the extent to which reflection on moral norms could also be a part of non-religious institutions that inculcate strong moral contemplation (e.g. de Botton, 2012), and help to build habits of virtue (e.g. Peysakhovich & Rand, 2015), or whether its content is specific to religious traditions that incorporate

beliefs, practices, and experiences oriented around supernatural and transcendent ideas. Exploring this in the context of ingroup favouritism would be particularly useful.

A third contribution of my work is to highlight that the extension of prosocial behaviour towards ingroups and outgroups is distinct from prejudicial attitudes extended towards the outgroup. In the DG, Christians who explicitly reported more negative attitudes towards atheists in fact still acted more prosocially towards atheists. Although mediation analysis indicated that this effect was explained by religiosity, rather than negative attitudes against atheists per se, it remains an interesting example of an intergroup context where explicit negative attitudes do not correspond with behavioural discrimination, suggesting that future work on anti-atheist and other forms of prejudice should take care to include behavioural manifestations of prejudice rather than reports of attitudes alone.

Limitations

Certain limitations of the present work bear mention (notwithstanding the potential limitation of my reliance on an MTurk sample, which I shall discuss more in the final chapter of this thesis). A first concern relates to demand characteristics. To ensure that participants attended to the group membership, it was salient in my design that the recipient was either a self-identifying atheist or Christian. This saliency of the recipient's belief may have introduced demand characteristics which could have affected results. Although I used a between-subjects design to reduce this potential (if a participant only completes a single trial, it is harder for them to guess the purpose), residual demand characteristics remain a possibility. Of course that said, it is difficult to know in advance

what precisely the 'demand' might have been and so regardless of my results I could say there is a concern with demand characteristics. For example, because participants saw the group membership of the other player, one could either say that there could be demand characteristics to act in an ingroup-favouring way (which is not what I saw, at least for religious participants), or that there could be a kind of second-order demand characteristics such that participants expected me to expect them to act in an ingroup favouring way and then deliberately acted in a different way. Moreover, if demand characteristics led Christians but not atheists to act impartially, this remains a theoretically interesting finding. If participants did guess that the study was about religion, and only Christians – but not atheists – were motivated to present their group in a positive light by not acting in a discriminatory way, this would be interesting in itself because atheist participants are the ones for whom acting prosocially would help improve their group's image and so may be especially motivated to respond to the demand characteristics in question. Overall, while demand characteristics are a possible (as, indeed, they are for most studies using economic games), I do not think this a damning problem.

A second concern is that I could have inadvertently decreased the salience of group membership by including identifying information. Recall that, in an attempt to reduce demand characteristics by making it too obvious that the phenomenon of interest was religious identity, I also provided participants with additional information about the recipient (i.e. living in the United States, speaking English as their primary language, and being over the age of 18). Scholars working on prejudice reduction and intergroup contact

have argued that one way to reduce intergroup bias is through *decategorization:* thinking of people less in terms of group identities and more in terms of personal identities. Put simply, the idea is that if group members can be encouraged to think of themselves as separate individuals rather than exemplars of their group (Wilder, 1981) and/or have personalized interactions with outgroup members to know them better as individuals (Pettigrew, 1997, 1998), through such decategorization intergroup bias will be reduced (Brewer & Miller, 1984; Ensari & Miller, 2001; S. L. Gaertner et al., 2000). Moreover, there is some evidence that focusing on individuating characteristics is more effective at reducing intergroup bias for majority than minority group members (Bettencourt, Charlton, & Kernahan, 1997). It is possible, therefore, that by including some identifying information I actually decreased the salience of the group memberships, and this is why I observed no ingroup favouritism on the part of Christians, who are a majority group in the US. That said, this potential concern should not be overstated. The 'individuating' information I provide was actually explicitly chosen to be as a neutral and non-descriptive as possible: it is a legal requirement that MTurk workers are over the age of 18; for workers to take part in tasks they must speak English because this is the language the tasks and website are in; and MTurk studies in psychology typically only use American participants. In contrast, the kind of individuation focused on in decategorization is typically much more specific to the person – their personal background, preferences, desires, etc. Thus, the degree to which this is really individuating information is open to question. Nonetheless, this remains a potential limitation and should be kept in mind.

A third problem with these studies concerns the change I made to the DG to make it comparable to the TG: specifically, having it so that transfers in the DG were doubled by the experimenter. This does not change the basic idea of the game – giving money at economic cost to the self with no potential for reciprocity or reputation-formation – but it does introduce the potentially confounding factor of efficiency because even transferring a small amount (e.g. 5%) would give the recipient more at no additional cost to the participant (e.g. 10%). It is possible, then, that religious people transferred more than atheist participants because they were more concerned with economic efficiency. That said, this is not a critical problem because this element of efficiency did not change the key feature I was interested in – transferring more of one's own money to benefit another. Even if decisions were driven by efficiency concerns, this does not change the fact that the behaviour is still prosocial as defined in this thesis. Moreover, even if prosocial behaviour was driven primarily by efficiency, it remains theoretically interesting if religious people were more concerned by efficiency than atheists, and if efficiency led atheists but not Christians to show ingroup favouritism. An appeal to efficiency, then, is really just a re-specification of the question 'Do people show ingroup-favouring or generalized prosociality?' into the question 'Do people show ingroup-favouring or generalized concerns about the efficiency of monetary outcomes to others?'. Nonetheless, it would be useful for future work to use a standard, non-doubled, DG to address this concern.

A final limitation concerns the degree of religiosity in my Christian participants. While the Christian participants in the sample were moderately religious, it remains to

be seen whether these results would be observed amongst participants who are very highly religious and enmeshed in their religious communities. Such individuals would likely place more importance on their religious social identity, but the extent to which this would increase prosociality in general and/or prosociality towards co-religionists specifically remains unclear. It would be useful, therefore, for future work to explore the extent of religious prosociality in less artificial contexts and with more participants who are extremely high on religiosity.

Conclusion

In conclusion, in this final empirical chapter I sought to explore ingroup favouritism in the context of religious groups and found intriguing results whereby atheists but not Christians showed ingroup favouritism in an anonymous DG with no interdependence of outcomes. In doing this, I made further contributions to our understanding of religious prosociality. In my next chapter I wanted to follow this up by looking again at comparing behaviour in the DG and TG, but this time looking at behaviour extended towards negatively-stereotyped outgroups.

Chapter 5: General Discussion

Summary of Research Findings

In this thesis, I have reported three empirical chapters looking at ingroup-favouring prosocial behaviour. Before discussing the contributions of this thesis in the next section, I shall briefly recap my three empirical chapters and the findings therein.

In Chapter 2 I looked at decision framing and whether mere categorization can be sufficient to induce ingroup favouritism. Using a newly-developed task – the Intergroup Lottery Task (ILT) – I had participants decide, across a series of trials with different levels of economic cost and efficiency, whether they would pay some of their own money in the task to give another player a better chance of earning more money. I conducted this first in a gains frame (Study 1) and then in a losses frame (Study 2), and manipulated within-subjects across trials whether decisions were public or private, and whether the other player was an ingroup or outgroup member. My results showed that mere categorization was sufficient to induce ingroup favouritism: minimal groups were significantly more prosocial to ingroup members than outgroup members, and that this was the case regardless of whether decisions were public or private, regardless of the economic cost of helping, and regardless of gains and losses frames.

In Chapter 3, I explored the intuitive basis of ingroup favouritism. My results from Chapter 2 suggested that ingroup favouring prosocial behaviour – like prosocial behaviour more generally - might have roots in intuitive, automatic processes. In Chapter 3 I followed up on this idea and applied dual-process models to intergroup prosocial behaviour. A growing body of work has demonstrated intuitive social preferences whereby people are 'intuitive cooperators' (Rand, 2016; Zaki & Mitchell,

2013). And yet at a theoretical level, the idea that intuition uniformly favours cooperation—even with outgroup members—is at odds with much past evidence illustrating implicit outgroup bias and distrust (Cikara & Fiske, 2012; Correll et al., 2002). I therefore conducted a rigorous pre-registered test of intuitive prosociality that addressed methodological problems that exist in the literature. Participants completed a Prisoner's Dilemma (PD) game, and I manipulated whether they were encouraged to make their decision either intuitively, through deliberation, or in the absence of either instruction; the group membership of the other player; and whether the PD was framed in terms of gains or losses. My results showed that participants were less cooperative under conditions of reflection than intuition, more cooperative in gains frames, and more cooperative with ingroup than outgroup members. And importantly, while intuition increased prosociality generally, it failed to eliminate bias: even under time pressure participants cooperated more with arbitrary ingroup than out members.

In Chapter 4, I explored ingroup-favouring prosocial behaviour in the context of real groups, and looked at whether there are differences in patterns of ingroup favouritism in non-interdependent contexts where there is no potential for reciprocity (DG) compared to contexts where decisions are interdependent and expectations of reciprocity are critical (TG). I looked at how much Christian and atheist participants transferred in a DG and TG to either Christian or atheist recipients. In both the DG and TG, Christian participants were more prosocial than atheist participants, and in the DG atheist participants, but not Christians, showed ingroup favouritism. My results showed that Christian participants transferred more both to Christian and atheist recipients, and

this was driven by the frequency of thinking about religious ideas. Interestingly, self-reported negative attitudes on the part of Christians towards atheists did not translate into actual behavioural discrimination.

Contributions

What have we learned? As outlined in each empirical chapter, my work has made numerous contributions to understanding of prosocial behaviour in general, including how gains and losses frames influence prosocial behaviour (Chapters 2 and 3), whether selection bias accounts for the finding that time pressure increases cooperation (Chapter 3), and whether religious people really are more prosocial (Chapter 4). But moreover, and more critically for my thesis, my work has advanced our understanding of ingroup-favouring prosocial behaviour.

First, I have shown that while whether a decision is framed in terms of gains or losses has a significant influence on prosocial behaviour generally, ingroup favouritism specifically seems relatively unaffected by such framing effects. Second, I have shown that despite claims to the contrary, simply sharing a social identity can be sufficient to induce ingroup favouritism even in the absence of interdependence and expectations of reciprocity. Third, I have shown that ingroup favouritism in prosocial behaviour has its roots in intuitive processes, being observed even without the potential for deliberation. Together, these contributions paint a picture by which acting more prosocially towards ingroup than outgroup members constitutes a 'default' position based on intuitive process that are activated across different contexts and even with novel groups.

Ingroup Favouritism in Gains and Losses

The first contribution of my work reported in this thesis is a novel examination of ingroup-favouring prosocial behaviour in both gains and losses frames (Kahneman & Tversky, 1979). That whether a choice is framed in terms of gains or losses can lead to divergent decisions is well-documented, with most research showing the phenomenon of loss aversion whereby a loss has a stronger influence on choices than a gain of the same magnitude (Kahneman & Tversky, 1979; for a meta-analysis, see Kühberger, 1998). However, when it comes to loss aversion in prosocial contexts, the evidence has been mixed, and there has been almost no work at all on loss aversion in ingroup-favouring prosocial behaviour. To meet this gap in the literature, in my first two chapters I looked at loss aversion in ingroup favouritism and found a pattern of results that was simultaneously convergent and divergent across the two chapters.

On one hand, both chapters showed that ingroup favouritism was observed in both gains and losses frames, consistent with the idea that ingroup favouritism is a robust phenomenon that potentially runs deep: even when the situation is framed in very different ways, people tend to favour their ingroup over their outgroup. This suggests that, while research on ingroup favouritism has almost exclusively focused on gains frames and, to my understanding, never compared ingroup favouring prosocial behaviour in both gains and losses contexts, this may not be as much of a problem as might first appear because these frames do not seem to have a strong influence on ingroup favouritism.

On the other hand, while both chapters showed that framing influenced the incidence of non-ingroup-specific prosocial behaviour, the direction of how this framing influenced prosocial behaviour was different in the two chapters such that there was more prosocial behaviour observed in the losses frame in Chapter 2 and more prosocial behaviour observed in the gains frame in Chapter 3. As I have outlined in my discussion of Chapter 3, one possibility is that the loss-framed PD in Chapter 3 was more cognitively complex than the loss-framed ILT in Chapter 2, both because of the interdependent structure and the mental calculations involved. It would be interesting for future research to explore this possibility.

It is interesting, then, that overall and contrary to previous research, losses and gains do not seem to make much difference in an intergroup context. This seems to support arguments made by Li and colleagues (2012), who argue that loss aversion represents a "domain-specific bias operating according to an adaptive logic of recurring threats and opportunities in different evolutionary domains" (p.550). Loss aversion, in other words, may be manifested in different ways and in different strengths in different contexts, and one context in which loss aversion may be less potent is in intergroup encounters – potentially because here, other more salient intergroup-focused decision biases and inclinations outweigh the effects of loss aversion. A great deal of research has been conducted on the contexts in which loss aversion is observed. One implication of my work is that it would be interesting to also look more at the contexts in which loss aversion is less important.

'Mere Categorization' vs. Reputation and Reciprocity

The second contribution of my work is to highlight that simply sharing a social identity – 'mere categorization' – can be sufficient to cause ingroup-favouring prosocial behaviour. As outlined in the introduction, there are two general categories of explanations of ingroup-favouring prosocial behaviour, with some accounts suggesting that simply sharing a group membership can induce ingroup favouritism through social identity processes like depersonalization (Tajfel & Turner, 1979; Turner et al., 1987) (Brewer, 1979; De Cremer & Van Vugt, 1999; Simpson, 2006; Tajfel & Turner, 1979; Turner et al., 1987), and others suggesting that people show ingroup favouritism in prosocial behavior only because they expect favourable treatment from fellow ingroup members (Rabbie et al., 1989; Yamagishi & Kiyonari, 2000; Yamagishi & Mifune, 2008). The critical test between those approaches for whether social identity can be sufficient to induce ingroup favouritism in minimal groups comes when decisions are private and so the participant cannot be identified as an ingroup member by the recipient.

In Chapter 2, I explored this and found (amongst other things) that simply sharing a social identity was enough to induce ingroup favouritism: ingroup-favouring prosocial behaviour was observed even when there was no interdependence and potential for reciprocity, and even when decisions were completely anonymous and there was no potential for reputation-building. Moreover, participants' degree of ingroup favouritism (i.e. helping ingroup members more than outgroup members) was relatively distinct and uncorrelated with the degree to which they exhibited reputation-based helping (i.e. helping more in public than private conditions), and that participants who showed

ingroup favouritism in public conditions were also more likely to do so in private conditions. These results highlight that mere categorization can be enough to induce ingroup favouritism, seemingly through activating social identity-based heuristics to help ingroups more than outgroups. This, of course, is inconsistent with the claims of the bounded generalized reciprocity (BGR) model, which posits that "in-group favouritism *does not* occur when participants cannot or do not expect favourable treatments from in-group members" (Yamagishi & Kiyonari, 2000, p. 127, my italics). Therefore, while there is a lot that the BGR models get right – for example, the focus on the general importance of reciprocity in intergroup contexts –the claim that reciprocity and reputation-building is the entirety of ingroup favouritism is too strong. Rather, as suggested by Brewer (2008) there can be two routes to ingroup-favouring prosocial behaviour such that both mere categorization social identity processes and reciprocity can play a fundamental role.

Intuitive Ingroup Favouritism

My results paint a picture by which ingroup favouritism results from fairly intuitive, automatic processes. In Chapter 2 I showed that that ingroup favouritism was observed regardless of the economic cost and benefit of helping, and that participants who showed ingroup favouritism in public conditions were also more likely to do so in private conditions. This is suggestive of ingroup favouritism being driven by a fairly intuitive and automatic application of a social heuristic to help ingroup members more than outgroup members, rather than any deliberative and controlled weighing up of the utility to be gained from helping an ingroup versus an outgroup member or the strategic benefits associated with this. Similarly, in Chapter 3, I showed that even under time

pressure where the ability to deliberate is reduced, people help ingroup members more than outgroup members. Moreover, that ingroup favouritism was observed even in an unconstrained control condition highlights that cooperating with the ingroup more than the outgroup is a 'default' stance in human cognition.

Of course this does not mean that more controlled, deliberative processes have no input into ingroup favouritism. Rather, what my results suggest is that intuitive processes do play a critical role such that ingroup favouritism will likely be observed as a default unless there is an opportunity and motivation to 'correct' this behaviour through more deliberative processes. Exploring what factors – for example, reflection on religious and moral norms - may lead to an individual being more likely to correct the intuitive preference for the ingroup and instead act in a more impartial way will be an interesting direction for future research.

Limitations

MTurk

One aspect of my studies that requires further discussion is my use of MTurk participants in Chapters 3 and 4. To recap, Amazon Mechanical Turk (MTurk) is an online labour market powered by Amazon in which participants ("workers") can complete online tasks ("HITs") posted by researchers ("requesters") for money. Typically, these studies are short (e.g. less than 10 minutes) and the payment is low (e.g. less than $2.00). After participants complete the task, the researcher then approves (or much more rarely,

rejects[41]) their work and the participant is paid. According to Amazon, the MTurk website boasts a global subject pool consisting of thousands of participants ready to complete thousands of tasks at any time of any day. Correspondingly, MTurk allows researchers to obtain large number of participants in a short time frame for a fraction of the cost that it would usually cost to recruit in a physical laboratory. This is seen, for example, in my work reported in Chapter 3, where it would have been almost impossible to recruit and pay over 1,000 participants to come into the lab and complete the task. MTurk represents a convenient source of data, but it is a little *too* convenient?

While there are few published papers explicitly arguing against the validity of MTurk samples, concerns have been raised about whether the quality of MTurk data is comparable to that obtained from traditional lab-based student samples used in psychology. It is suggested, for example, that because the levels of compensation are low and the implementation is online, MTurk participants can be inattentive, fail to follow instructions, or even deliberately mislead the experimenter. This, of course, would mean that the quality of MTurk data would be lower than that from student-based samples.
In this section I will outline some of the strengths and weakness of conducting studies online and argue that while there are certain concerns and issues that researchers must be aware of when using MTurk, overall it remains a reliable and efficient method of collecting data.

---

[41] This rejecting of workers is, to my understanding, rare in psychology due to research ethics dictating that one pays the subject even if they decide to opt out of the study after completing. Correspondingly, I have never rejected and not paid a worker on MTurk, though I have excluded them from my data analysis.

Problems with MTurk

Attentiveness

One prominent criticism of the use of MTurk revolves around the idea that MTurk participants may be inattentive and not respond to questions properly, potentially because participants receive low compensation for taking part and/or because the online format affords anonymity. It is widely accepted – even by enthusiasts of MTurk – that some participants do not pay attention, and this has led to the widespread adoption by researchers of 'attention checks' to see which participants are not paying attention (often, but not always, such participants are excluded from data analysis). There are at least two closely related forms of such measures: first, so-called instructional manipulation checks where participants are given a 'trick' question (IMC: Oppenheimer, Meyvis, & Davidenko, 2009); and second, comprehension checks where participants are asked to relay some information about the nature of the task. While attention checks are widely used in online studies – including mine here[42] - just the fact that some participants do not pay attention and that researchers have to use these attention checks to filter out 'bad' participants is not itself proof that MTurk data is of a lower quality than traditional student samples. To do this, one must directly compare MTurk samples with student samples – and luckily there are a few studies doing this. At least one study shows that MTurk participants are more likely to fail IMCs than student samples (Study 2,

[42] For example, in Chapter 3 I had forced comprehension questions whereby participants had to answer questions about the PD correctly before they could continue, and post-task comprehension questions whereby after completed the PD I asked participants to report the group membership of their partner so that I could exclude from analysis anyone who got this wrong.

Goodman, Cryder, & Cheema, 2013). A larger body of research findings, however, challenges this idea that MTurk participants are more inattentive than student samples. Some studies have found that MTurk participants do not differ from community samples in how much they pass or fail IMCs (Study 1, Goodman et al., 2013; Paolacci, Chandler, & Ipeirotis, 2010). Moreover, some studies have found that MTurk participants are actually less likely to fail IMCs than student samples, indicating they are more attentive. For example, a "Many Labs" replication project looked at several effects in psychology (e.g. gains vs. loss framing; anchoring; imagined contact) across different countries and methods of implementation and in this process found that MTurk participants passed IMCs at a higher rate than student samples (Klein et al., 2014). And most importantly, Hauser and Schwarz (2016) directly compare the attentiveness of MTurk and student samples and find across three studies that MTurk participants were more likely to pass even novel IMCs.

Why this change from earlier studies suggesting MTurk participants be less attentive? One possibility is that MTurk's in-built reputation system helps ensure a reasonable degree of attentiveness. Requesters can reject submissions of workers if they have not completed the study properly, and many researchers (like I have done in my studies here) restrict participation to workers who have only a small percentage of their work rejected, creating an incentive for MTurk participants to complete work properly. Consequently, the MTurk population may have learned to pay more attention to instructions and questions in order not to rejected and thus lose out of future earnings

(Hauser & Schwarz, 2016), which is consistent with findings that that workers with high reputation scores are more likely to pass IMCs (Peer, Vosgerau, & Acquisti, 2014).

Further evidence against the idea that MTurk samples are less attentive than traditional sample size comes from considering the claim that the low payment makes people less invested and motivated to complete the study properly. If the small monetary pay makes the data quality lower, we would expect that there be higher quality of data when the pay was higher. In fact, this does not seem to be the case. For example, Buhrmester, Kwang, and Gosling recruited MTurk participants in a 3x3 design in which they varied the monetary compensation ($0.02, $0.10, or $0.50) and the task completion time (5, 10, or 30 minutes) and gave participants six personality questionnaires. Looking at reliability analyses of the scores, they found that "even at low compensation rates, payment levels do not appear to affect data quality; the only drawback appears to be data collection speed" (Buhrmester et al., 2011, p. 4). That is, paying participants only 2 cents for a 10 minute task might make data collection slower, but it doesn't seem to affect data quality. The problem with this study for my purposes, of course, is that even the highest payment rate offered ($0.50) is still far below what participants would be paid in a laboratory context and below the US minimum wage[43]. A better test of how pay (and the online format with anonymity) affects data quality comes from work by Casler, Bickel, and Hackett (2013). They compared the quality of data in a short 10-minute object categorization task obtained from traditional face-to-face behavioural testing with that

[43] For a 5 minute study, the current US minimum wage of $7.25 per hour would dictate a minimum payment of $0.60, which is above the payment rates here.

from MTurk. In the lab-based version, participants were paid $5.00 and were given information by an experimenter about various tools face-to-face. In the MTurk version, participants were paid $0.50 and watched a video of the experimenter's face-to-face instructions. Again, they found that the quality of the data and the empirical results were virtually identical across the two conditions, suggesting that whether an experiment is conducted face-to-face or online and whether the payment is large or small[44] does not have much impact on the quality of data.

Misrepresentation

Some recent work has suggested that one problem with conducting studies online is that MTurk workers can misrepresent themselves by claiming a false identity or activity in order to qualify, and thus get paid, for a study (Sharpe Wessling, Huber, & Netzer, 2017). This is a problem where in the description of the survey posted on MTurk or on the first page of the task, the researchers say that the study is only open to people who fulfil some criteria – for example, being of a certain ethnic background, being religious or owning a car[45]. Because workers want to complete the study, there is evidence that a large number of participants will misrepresent themselves and claim to fulfil this criteria when they do not. And particularly worryingly, workers who misrepresent themselves in this way respond in substantially different ways to appropriately identified participants (Sharpe Wessling et al., 2017). Because in my studies I just excluded from

[44] Of course, "large" or "small" also depends on the context: a payment of $1.50 would be very high for MTurk, but very low for a traditional lab-based study.

[45] This is different from a self-enhancement bias whereby participants are motivated to be seen in a positive way and therefore might misrepresent themselves on socially desirable dimensions. This is likely to be just as much a problem – if not more – for student samples recruited in the laboratory.

analysis those who did not meet the criteria, there was no incentive to lie and in the lack of this economic incentive there is little identity misrepresentation (Sharpe Wessling et al., 2017).

Nonresponse and Selective Attrition

Another important problem with MTurk – as highlighted in my discussion in Chapter 3 – is that nonresponse and selective attrition can lead to a selection bias. Put simply, if in online studies certain types of participant are more likely to not answer questions (nonresponse bias: Wåhlberg & Poom, 2015) and/or drop out of the study (selective attrition: Zhou & Fishbach, 2016), this can lead to a selection bias that compromises random assignment and sampling and therefore confounds results. There are many reasons why participants might drop out, but one of the most commonly discussed is drop-out due to the complexity or length of the task. This might be a particular problem for online studies where drop out is less visible and less costly than in traditional laboratory-based studies[46]. Indeed, an analysis of selective attrition in online studies has suggested that dropout and nonresponse can be very high – approximately 30% to 50% of the sample (Zhou & Fishbach, 2016).

Selection bias can compromise conclusions of research. For example, as discussed in Chapter 3, a problem with previous studies looking at the effects of time pressure on cooperation is that they have typically excluded participants who failed to make their

[46] A student participant recruited to come into the laboratory might be less likely to drop out for such reasons, given that they've already 'committed', they've already made the effort to come to the testing place, and to drop out of a laboratory study would require telling the experimenter face-to-face, which for reasons like task complexity students might be unwilling to do.

decisions within the response window. These exclusion rates are typically very high and asymmetric across conditions – for example in Study 6 of Rand, Greene, and Nowak (2012), 48% of online participants failed to make their decisions under time pressure. That said, 46% of participants recruited in the laboratory for Study 7 also failed, suggesting that (at least in the context of time pressure studies) selection bias is not more critical for online than laboratory studies. Moreover, it seems that attrition and not answering questions is likely to be particularly problematic for long survey studies that involve open-ended text-based responses (Zhou & Fishbach, 2016). Because the studies used in my thesis are typically very short (< 10 minutes) and involve selecting pre-defined options rather than write long essay-type answers, it is not clear that this would have been a substantial problem for my work.

Strengths of MTurk

Notwithstanding the unique challenges that conducting studies on MTurk pose, there are also important strengths of this approach. First, MTurk is an efficient way of collecting data. As most researchers can attest, two of the biggest impediments to conducting research are limited time and limited money. Conducting studies in the laboratory with student samples can be much more expensive than using MTurk. For example, in my lab-based studies in Chapter 2 each participant was paid between £15 and £20, while in my MTurk studies in Chapter 4 each participant was paid less than $2.00. Similarly, while collecting the data for the studies in Chapter 2 required many hours over many weeks, the data on MTurk could be collected with little effort in a matter

of days. Given this, it is no surprise that MTurk has been such a popular way of collecting data in recent years.

Second, MTurk yields data that is more representative. There have been serious concerns about the reliance in psychology on student samples (Sears, 1986), and particularly the reliance on American students (Henrich, Heine, & Norenzayan, 2010). Participants from cultures that are Western, Educated, Industrialized, Rich, and Democratic (WEIRD) differ in myriad ways from participants who not WEIRD, and even within WEIRD cultures there are differences between students and non-students (Henrich et al., 2010). MTurk offers an effective way to help diminish this concern. For example, Buhrmester et al. (2011) looked at 3000 MTurk participants and found a roughly equal gender split (57% female), with around one-third being non-white (36%) and/or non-American (31%). Participants came from over 50 different countries and all 50 US states and included substantially more people of above-college age (M_{age}= 32). And while Casler, Bickel, and Hackett (2013) found no difference in the quality of data as indicated by the specific findings, they did find that the MTurk sample was significantly more diverse and representative than their lab-based sample. This increased diversity was particularly useful for me in Chapter 4, where I looked at ingroup favouritism amongst religious and atheist individuals. By using MTurk, I was able to recruit a very large number of participants from a variety of religious backgrounds and denominations. In contrast, because the UK in general - and student samples especially - are quite non-religious, it would have been extremely difficult for me to recruit enough actively religious people to come into the lab at Oxford. Moreover, of course, in many other ways

Oxford students are highly unrepresentative in terms of their ethnicity, education level, IQ, and class background. Thus, using MTurk likely afforded me a more representative sample than I could have obtained from using a traditional student sample.

Assessing the Quality of MTurk Data

Ultimately, when it comes to the quality of data obtained from MTurk, the proof is in the pudding. Many findings that have been demonstrated in lab-based studies have also been demonstrated in MTurk samples. For example, Horton, Rand, and Zeckhauser (2011) conduct three different studies on MTurk and find similar results to that obtained in the laboratory: in cooperation rates in a PD; the effect of gains/loss framing on medical decisions; and the effect of religious priming on cooperation in a PD. This, they conclude, is evidence that "online experiments can be just as valid— both internally and externally—as laboratory and field experiments, while requiring far less money and time to design and to conduct" (Horton et al., 2011, p. 399). MTurk yields comparable data quality to traditional lab tasks by replicating classic cognitive effects such as the Stroop effect and subliminal priming (Crump, McDonnell, & Gureckis, 2013). And in doing so, it is substantially cheaper, more time-efficient, and allows access to more representative samples.

In conclusion, it seems that MTurk is an effective platform for cheaply and quickly collecting data that is of a sufficient scientific quality and is more representative than that which can be obtained from traditional student samples alone (Buhrmester et al., 2011; Casler et al., 2013; Horton et al., 2011). That said, one should be careful not to entirely neglect lab-based samples. Time-intensive studies with multiple participants per testing

session (as in my work in Chapter 2) are unlikely to be best suited to online platforms like MTurk, let alone those looking at direct face-to-face interactions, and so there remains a vital role for traditional testing in social psychology. Ultimately, MTurk should be seen as an alternative to lab studies, not a replacement.

Deception

Another concern with the studies reported in my thesis is deception - "the intentional and explicit provision of erroneous information" in a study[47] (Hertwig & Ortmann, 2008, p. 222). Across the studies reported in this thesis I had varying levels of deception, ranging from almost no deception at all (Chapter 4) to more substantial deception where participants believed their decisions would affect another participant's payoff (Chapter 2), or where participants were told they were playing with another participant but were not (Chapters 3). Why this discrepancy? The pragmatic answer is that I ran these studies with different people at different research groups, and different research groups have different internal policies regarding deception. This difference in practices largely maps onto the distinction between attitudes towards deception in social psychology and behavioural economics. While in economics deception is all but forbidden and papers that use deception and are submitted to economic journals are extremely likely to be rejected, psychology takes a much more laissez-faire view. Having

[47] Deception refers to providing misleading information, rather than withholding information about, for example, the hypotheses or different experimental manipulations. As one economist who has argued against the use of deception in experiments has noted, "there is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not" (Hey, 1998, p. 397).

deception in these economic game tasks is substantially cheaper, easier, and more time-efficient, but do the costs of deception outweigh these benefits?

In economics, deception is not used because it is thought that doing so pollutes the common resource of trusting participants by causing participants in general to become distrustful of the experimenter, which in turn reduces data quality. To quote one particularly instructive passage from the economist John Ledyard:

> "It is believed by many undergraduates that psychologists are intentionally deceptive in most experiments. If undergraduates believe the same about economists, we have lost control. It is for this reason that modern experimental economists have been carefully nurturing a reputation for absolute honesty in all their experiments… if the data are to be valid, honesty in procedures is absolutely crucial. Any deception can be discovered and contaminate a subject pool not only for that experimenter but for others. Honesty is a methodological public good and deception is equivalent to not contributing."
>
> (Ledyard, 1995, p. 134)

As this quote suggests, economists often assume that psychologists have already polluted the pool by the frequent use of deception. The challenge for economists is largely to prevent this pollution crossing the disciplinary boundaries: "it is an unfortunate fact that experiments in psychology are tainted by distrust. We do not want the same taint to

be attached to experiments in economics" (Hey, 1991, p. 21). If psychology is already tainted, then, should one not just use deception as a matter of course?

While deception is much more acceptable in psychology, guidelines suggest that it should only be used as a last resort. According to the guidelines of the American Psychological Association, "Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's significant prospective scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible". The problem, of course, is that this "leaves the decision of whether or not deception is justified by its anticipated value to those who stand to benefit from its use" (Hertwig & Ortmann, 2008, p. 224). And indeed, many researchers do seem to think that deception is justified in many cases, because analyses on papers from the *Journal of Experimental Social Psychology* in 2002 and the *Journal of Personality and Social Psychology* in 1996 suggest that around half of all the articles included deception.

There are substantial benefits to using deception in studies using economic games, mainly involving efficiency in time and money. Consider, for example, my work in Chapter 2 where I recruited participants in groups of four for an hour-long session in which participants were typically paid around £15-20. I told participants that they were randomly selected to play as the decider in the game, and that two of the other participants were playing an entirely different game and would only receive money from the participant. This was deception that could, in principle, have been eliminated. I could have recruited participants to play a filler task and just be recipients, but this would

require me to recruit at least twice as many participants, taken at least twice as long and cost at least twice the price. And given that (as discussed in Chapter 2), it is particularly hard to recruit participants at Oxford and the research allowance afforded to D.Phil students quite low, this would likely have rendered the study impossible. And of course, given that participants typically already assume that psychologists routinely use deception[48] it is not clear that even using such procedures would make any difference to the quality of that specific data (participants wouldn't know they're being deceived until after they've completed the study) or how much impact it would make in un-polluting an already polluted field. In such cases, the benefits of deception may well outweigh the costs. That said, because deception does have the potential to cause problems, in future work I plan to adopt a no-deception practice wherever possible.

Directions for Future Research

Intuitive Intergroup Aggression

Throughout this thesis, I have focused on ingroup favouritism in prosocial behaviour: helping "us" more than "them". In future work, I hope to explore the more destructive side of this tendency: when do we aggress against "them" to protect "us"? Some scholars have suggested that ingroup-favouring prosocial behaviour has a strong link to intergroup competition, such that humans can be described as exhibiting *parochial altruism*: the two-fold tendency for a) increased cooperation and prosocial behavior

[48] A contention supported by the fairly large number of participants I had to exclude from the two studies in Chapter 2 for, amongst other reasons, disbelieving the experimental paradigm as described to them was real.

within the boundaries of a group and b) the rejection, derogation, and even harm of outgroup members (Bernhard, Fischbacher, & Fehr, 2006; Choi & Bowles, 2007; De Dreu et al., 2016). While at first blush these two components – one more about helping, the other more about aggression – might seem in conflict, evidence from De Dreu and colleagues suggests that in fact they are two sides of the same coin (De Dreu, Balliet, D., Halevy, N., & Elliot, A.J., 2014): The same mechanisms that have allowed for intragroup cooperation, on some accounts, have allowed for intergroup competition, and both were necessary for survival (Bernhard et al., 2006; Choi & Bowles, 2007; Rusch, 2014). I hope to explore this aspect of intergroup behaviour by looking at defensive aggression, or *pre-emptive striking*.

The idea of a pre-emptive war – where one launches war to gain a strategic advantage in an impending (and allegedly unavoidable) conflict shortly before that attack materializes – is present throughout history and central to contemporary discussions of the ethicality and legality of different kinds of war. Yet pre-emptive strikes are not confined to international contexts – they also occur at the individual, psychological, level. In his classic essay, Schelling (1960) uses the example of an armed homeowner who is faced with a burglar that also has a gun. If the homeowner believes that the burglar might shoot, then the best response is for her to shoot first in a surprise attack and incapacitate the burglar before he can hurt her. But suspecting this might happen, the burglar might also be tempted to shoot the homeowner first to incapacitate her before she can hurt him. In this way, both sides can end up attacking, even though neither had any desire to hurt the other. Indeed, evidence suggests that individuals do engage in pre-emptive striking:

"the simple fear of nastiness" is enough to make people inflict damage on another person, even when there is no economic benefit of doing so (Abbink & de Haan, 2014, p. 195)

Fear and (mis)trust appear to be at the very core of pre-emptive striking - and given that fear and trust are also critical drivers of intergroup conflict (Insko & Schopler, 1998; Sherif, 1966), an obvious question arises: are pre-emptive strikes more common in intergroup contexts when playing with some from the outgroup or ingroup member? Existing evidence is inconclusive: some work suggests that the motivation to protect the ingroup promotes pre-emptive aggression (Böhm, Rusch, & Gürerk, 2016) and that participants don't just care for the ingroup but are willing to defend when necessary (De Dreu et al., 2010), but other work suggests that in a minimal group paradigm participants are equally like to attack an ingroup and outgroup member (Simunovic, Mifune, & Yamagishi, 2013)

In future work, I hope to explore pre-emptive striking to help resolve two questions currently unaddressed in the existing literature. First, I plan on looking at pre-emptive striking in real intergroup contexts. Simunovic et al. (2013) found no effect of group membership, but it is possible that this was due to their use of minimal groups, which – by definition – lack a history of continued intergroup conflict where fear and distrust will be heightened. Second, I hope to apply the method of Chapter 3 and look at pre-emptive striking under conditions of time pressure or time delay. In doing this, I can explore whether intergroup pre-emptive striking is driven more by intuitive or more reflective, deliberative processes. Indeed, with the authors I collaborated with for my work in Chapter 3, I have already begun developing a Pre-Emptive Strike Game (PSG)

that can be administered under time pressure or reflection and preliminary data using real political groups suggests that when participants are placed under time pressure they pre-emptively strike against both ingroup and outgroup members equally. When given the potential to deliberate, however, participants are significantly less likely to aggress against an outgroup member. In future work, I hope to replicate and extend this preliminary finding by adding in a baseline condition where no group membership is specified to see whether deliberation is making participants selectively less likely to aggress against ingroup members, or is maintaining the tendency to aggress against outgroup members.

Parochial Morality

Morality is a central – even defining – feature of what makes humans who we are, and we develop intricate systems of social, legal, and moral norms that dictate how we should act, and condemn and punish those who do not comply. At the core of human morality is prosociality - helping, caring, and treating other people fairly – but of course our prosociality is not extended impartially. Rather, we exhibit well-documented tendencies for ingroup favouritism and intergroup competition that appear at odds with the universalistic moral rules that people adhere to and hold others accountable for (e.g. "thou shall not kill"; "treat others as you'd wish to be treated yourself"). Not only are we more likely to act in a prosocial – that is, moral - way towards people with whom we share a group membership, but these group memberships fundamentally shape what we think a moral act is. As well as demonstrating parochial altruism, we might be said to show parochial morality.

Just like research has highlighted a central role for intuitive, automatic cognitive processes in parochial altruism, other research has shown that intuitive moral judgments seem to track with the two ethical approaches of *deontology* (Kant, 1797/2002; Scanlon, 1998) and *consequentialism* (Bentham, 1789/1983; Mill, 1863). Consequentialist approaches like *utilitarianism* say we should aim to impartially maximize the greatest good for the greatest number, even if this means causing some harm – in the classic 'trolley problem', killing one person by pushing them off a bridge so that their body will stop the train and save the five people stuck on the track. In contrast, deontological approaches focus on moral rules and ideas of rights and duties, such that certain things (like killing an innocent person) are wrong even if they maximize good outcomes (like saving extra lives). A wealth of behavioral and neurobiological evidence has shown that participants' intuitive judgments tend to be more deontological whereas characteristically consequentialist judgments are often the result of slow, deliberative cognitive processes (Greene, 2007; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Koenigs et al., 2007). But why would this be? Almost by definition, wholesale adoption of utilitarian principles would lead to the greatest happiness for the greatest number of people, and consequently it might be surprising that we have deontological moral intuitions when these appear to lead to socially inferior outcomes. In some of my other work I have already begun working on this problem, looking specifically at why moral intuitions typically follow deontological prescriptions as opposed to those of other ethical theories by drawing on *mutualistic partner choice models* in evolutionary theory (Alexander, 1987; Baumard, André, & Sperber, 2013; Noë & Hammerstein, 1994).

Partner choice models posit a cooperation market such that agents who can be relied upon to act in a mutually beneficial way are more likely to be chosen as cooperation partners, thus increasing their own fitness (Alexander, 1987; Baumard et al., 2013; Krebs, 2008; Noë & Hammerstein, 1994; Trivers, 1971). People tend to select the most cooperative individuals as partners, and those who contribute less than others are gradually left out of cooperative exchanges (e.g. Barclay, 2004, 2006; Rockenbach & Milinski, 2011). Thus, if people who make certain types of moral judgments are favored in a cooperative market (potentially because these judgments signal a commitment to cooperation), so too will these judgments come to be favored as defaults (Baumard et al., 2013). Indeed, across 9 experiments with more than 2,400 participants, I found that people who took an absolute approach to the dilemmas (refusing to kill an innocent person, even when this maximized the greater good) were seen as more trustworthy than those who advocated a more flexible, consequentialist approach (Everett, Pizarro, & Crockett, 2016). In particular, this seemed driven by the fact that deontological judgments – compared to utilitarian consequentialist ones – typically reference the rights of other people and honour the implicit contractual obligations we have to one another.

In future work I would like to explore the connection between parochial altruism and the development of consequentialist and deontological moral intuitions. Specifically, I want to explore the role of deontological vs. consequentialist moral judgments (henceforth simply "moral judgments") in constructively regulating both the cooperative and competitive aspects of parochial altruism. This is an issue that has not yet been addressed, and yet there are important connections between the two phenomena that

suggest investigating the relationship between parochial altruism and moral intuitions will be critical. First, they share a common aetiology: both parochial altruism (Bernhard et al., 2006; Choi & Bowles, 2007; Rusch, 2014) and deontological moral intuitions (Everett et al., 2016) can be understood as consequences of the central evolutionary problem of facilitating cooperation with people around us to ensure survival. Second, they share common psychological manifestations, with more intuitive processes underpinning both parochial altruism (Correll et al., 2002; De Dreu et al., 2015) and deontological judgments (Greene et al., 2008). Third, both deontological moral intuitions and parochial altruistic tendencies seem to converge in seeing prosocial behaviour as typically being more important towards certain people than others. Given the strong link between the two areas of research, in future work I hope to explore this further in order to integrate and extend our understanding of the origins, processes, and consequences of parochial altruism and moral intuitions in interplay.

To reference just one approach that I hope to take on this topic, I hope to conduct explore the relationship between ingroup favouritism and moral intuitions by adapting the design of my previous work to an intergroup context. I will, like in Chapter 4, use a TG in which participants will play the role of the investor and be told that their partner is either an ingroup or outgroup member, and made either a deontological or consequentialist decision in a trolley-style moral dilemma where one must decide whether to sacrifice one person to save the lives of a greater number. Based on my previous findings, I predict that participants will be more likely to trust both an ingroup and outgroup member more when the partner expresses a deontological rather than a

consequentialist judgment, but because of greater moral norms about respect and rule-following within the boundaries of groups, that this difference will be stronger for an ingroup than an outgroup member. In follow-up work I would like to explore whether deontological participants more trustworthy and whether is this dependent on whether the other person is from an ingroup or outgroup.

Concluding Remarks

It would be amiss to consider the issue of ingroup favouritism as just a game that scientists play, with merely theoretical and methodological implications of interest just to other scientists. Rather, this issue is one that has far-reaching implications for society at large. The ability of *homo sapiens* to cooperate and act prosocially is undoubtedly one of the biggest – if the not *the* biggest - reasons why our species has flourished. A physically rather weak species, we have colonized the earth through cooperation and team work: the development of liberal democracies, the invention of the computer and the creation of life-saving drugs all arise from the ability of our species to work together. Despite this, however, it is clear that our cooperative tendencies still leave a lot to be desired. In a globalized world, our parochially altruistic tendencies with their roots in our ancestral past in the African savannah are no longer adequate. To ensure the survival of our species in the face of drastic and rapid climate change, poverty, population increases, famine, war, and disease we need now – more than ever – to work together across nations, ethnic groups, religious groups, and so on. Psychology has a very real and important role to play in exploring how we can encourage widespread cooperation across group lines. As

Martin Luther King, Jr said, "We must learn to live together as brothers or perish together as fools." (Speech in St. Louis, Missouri, March 22, 1964).

References

Abbink, K., & de Haan, T. (2014). Trust on the brink of Armageddon: The first-strike game. *European Economic Review*, *67*, 190–196. https://doi.org/10.1016/j.euroecorev.2014.01.009

Ahmed, A. M. (2007). Group identity, social distance and intergroup bias. *Journal of Economic Psychology*, *28*(3), 324–337. https://doi.org/10.1016/j.joep.2007.01.007

Ahmed, A. M. (2009). Are religious people more prosocial? A quasi-experimental study with madrasah pupils in a rural community in India. *Journal for the Scientific Study of Religion*, *48*(2), 368–374. https://doi.org/10.1111/j.1468-5906.2009.01452.x

Ahmed, A. M., & Salas, O. (2009). Is the hand of God involved in human cooperation? *International Journal of Social Economics*, *36*(1), 70–80. https://doi.org/10.1108/03068290910921190

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881

Alexander, R. (1987). *The biology of moral systems*. New York, NY: Aldine de Gruyter.

Allport, G. W. (1954). *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley Pub. Co.

Alvard, M. (2001). Mutualistic hunting. In C. Stanford & H. Bunn (Eds.), *Meat-Eating and Human Evolution* (pp. 261–278). Oxford, England: Oxford University Press. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=4xFnDAAAQBAJ&oi=fnd&pg=PA261&ots=gXhuAeF2LN&sig=8VZwKuhHzP-XvwbqYeWdsATnwbA

Anderson, L. R., & Mellor, J. M. (2009). Religion and cooperation in a public goods experiment. *Economics Letters*, *105*(1), 58–60. https://doi.org/10.1016/j.econlet.2009.05.016

Anderson, L. R., Mellor, J., & Milyo, J. (2010). Did the devil make them do it? The effects of religion in public goods and trust games. *Kyklos*, *63*(2), 163–175. https://doi.org/10.1111/j.1467-6435.2010.00456.x

Artavia-Mora, L., Bedi, A. S., & Rieger, M. (2017). Intuitive help and punishment in the field. *European Economic Review*, *92*, 133–145. https://doi.org/10.1016/j.euroecorev.2016.12.007

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: articulation and significance of multidimensionality. *Psychological Bulletin*, *130*(1), 80–114. https://doi.org/10.1037/0033-2909.130.1.80

Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evolution and Human Behavior*, *32*(1), 41–49. https://doi.org/10.1016/j.evolhumbehav.2010.07.008

Axelrod, R. (1980). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(3), 379–403. https://doi.org/10.1177/002200278002400301

Bailey, R. C., & Doriot, P. D. (1985). Perceptions of professionals who express religious beliefs. *Social Behavior and Personality: An International Journal*, *13*(2), 167–170.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547. https://doi.org/10.1177/1368430209105040

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological Bulletin*, *140*(6), 1556–1581. https://doi.org/10.1037/a0037737

Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the "tragedy of the commons." *Evolution and Human Behavior*, *25*(4), 209–220. https://doi.org/10.1016/j.evolhumbehav.2004.04.002

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, *27*(5), 325–344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, *11*(2), 122–133. https://doi.org/10.1007/s10683-007-9172-2

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479. https://doi.org/10.1037/0003-066X.54.7.462

Barmettler, F., Fehr, E., & Zehnder, C. (2012). Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory. *Games and Economic Behavior*, *75*(1), 17–34. https://doi.org/10.1016/j.geb.2011.09.003

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, *8*(2), 71–83. https://doi.org/10.1002/bdm.3960080202

Baron, J. (1996). Do no harm. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics* (pp. 197–213). New York, NY: Russell Sage Foundation.

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.

Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78. https://doi.org/10.1017/S0140525X11002202

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, *113*(4), 936–941. https://doi.org/10.1073/pnas.1517780113

Bellemare, C., & Kröger, S. (2007). On representative social capital. *European Economic Review*, *51*(1), 183–202. https://doi.org/10.1016/j.euroecorev.2006.03.006

Bentham, J. (1983). *The collected works of Jeremy Bentham: Deontology, together with a table of the springs of action ; and the article on utilitarianism*. Oxford, England: Oxford University Press. (Original work published 1789)

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*(7105), 912–915. https://doi.org/10.1038/nature04981

Bettencourt, B. A., Charlton, K., & Kernahan, C. (1997). Numerical representation of groups in cooperative settings: social orientation effects on ingroup bias. *Journal of Experimental Social Psychology*, *33*(6), 630–659. https://doi.org/10.1006/jesp.1997.1334

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*(1), 27–52. https://doi.org/10.1002/ejsp.2420030103

Blanz, M., Mummendey, A., & Otten, S. (1997). Normative evaluations and frequency expectations regarding positive versus negative outcome allocations between groups. *European Journal of Social Psychology*, *27*(2), 165–176.

Böhm, R., Rusch, H., & Gürerk, Ö. (2016). What makes people go to war? Defensive intentions motivate retaliatory and preemptive intergroup aggression. *Evolution and Human Behavior*, *37*(1), 29–34. https://doi.org/10.1016/j.evolhumbehav.2015.06.005

Bontempo, R. N., Bottom, W. P., & Weber, E. U. (1997). Cross-Cultural Differences in Risk Perception: A Model-Based Approach. *Risk Analysis*, *17*(4), 479–488. https://doi.org/10.1111/j.1539-6924.1997.tb00888.x

Bourhis, R. Y., Sachdev, I., & Gagnon, A. (1994). Intergroup research with the Tajfel matrices: Methodological notes. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium, Vol. 7* (pp. 209–232). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., … Wollbrant, C. E. (2017). Registered Replication Report: Rand, Greene, and

Nowak (2012). *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691617693624

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

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

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B. (2004). Taking the social origins of human nature seriously: Toward a more imperialist social psychology. *Personality and Social Psychology Review*, *8*(2), 107–113. https://doi.org/10.1207/s15327957pspr0802_3

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 215–232). New York, NY, US: Psychology Press.

Brewer, M. B. (2011). Optimal distinctiveness theory: its history and development. In P. A. M. Van Lange, A. Kruglanski, & E. T. Higgins, *Handbook of theories of social psychology* (Vol. 2, pp. 81–98). London, England: SAGE. https://doi.org/10.4135/9781446249222.n30

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology*, *50*(3), 543–549. https://doi.org/10.1037/0022-3514.50.3.543

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. B. Brewer, *Groups in contact: The psychology of desegregation* (pp. 281–302). Orlando, FL: Academic Press.

Brooks, A. C. (2007). *Who really cares: The surprising truth about compassionate conservatism*. New York, NY: Basic Books.

Buhl, T. (1999). Positive-negative asymmetry in social discrimination: meta-analytical evidence. *Group Processes & Intergroup Relations*, *2*(1), 51–58. https://doi.org/10.1177/1368430299021004

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Bulbulia, J., & Mahoney, A. (2008). Religious solidarity: the hand grenade experiment. *Journal of Cognition and Culture*, *8*(3), 295–320. https://doi.org/10.1163/156853708X358191

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction* (Vol. xv). New York, NY: Russell Sage Foundation.

Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of human sociality* (pp. 55–95). New York, NY: Oxford University Press.

Campbell, D. T. (1965). Ethnocentric and other altruistic motives. In D. Levine, *Nebraska symposium on motivation* (pp. 283–301). Lincoln, Nebraska: University of Nebraska Press.

Caporael, L. R. (1997). The evolution of truly social cognition: the core configurations model. *Personality and Social Psychology Review*, *1*(4), 276–298. https://doi.org/10.1207/s15327957pspr0104_1

Capraro, V., & Cococcioni, G. (2016). Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports*, *6*, 27219. https://doi.org/10.1038/srep27219

Capraro, V., Jordan, J. J., & Rand, D. G. (2014). Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Scientific Reports*, *4*, 6790. https://doi.org/10.1038/srep06790

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Castano, E. (2008). On the perils of glorifying the in-group: intergroup violence, in-group glorification, and moral disengagement. *Social and Personality Psychology Compass*, *2*(1), 154–170. https://doi.org/10.1111/j.1751-9004.2007.00040.x

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology* (Vol. xiii). New York, NY, US: Guilford Press.

Chang, L. W., Krosch, A. R., & Cikara, M. (2016). Effects of intergroup threat on mind, brain, and behavior. *Current Opinion in Psychology*, *11*, 69–73. https://doi.org/10.1016/j.copsyc.2016.06.004

Chang, O. H., Nichols, D. R., & Schultz, J. J. (1987). Taxpayer attitudes toward tax audit risk. *Journal of Economic Psychology*, *8*(3), 299–309. https://doi.org/10.1016/0167-4870(87)90025-0

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior. *Journal of Political Economy*, *114*(3), 517–537. https://doi.org/10.1086/503550

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, *99*(1), 431–57. https://doi.org/10.1257/aer.99.1.431

Choi, J.-K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, *318*(5850), 636–640. https://doi.org/10.1126/science.1144237

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., & Fiske, S. T. (2013). Their pain, our pleasure: stereotype content and schadenfreude. *Annals of the New York Academy of Sciences*, *1299*(1), 52–59. https://doi.org/10.1111/nyas.12179

Cikara, M., Jenkins, A. C., Dufour, N., & Saxe, R. (2014). Reduced self-referential neural response during intergroup competition predicts competitor harm. *NeuroImage*, *96*(1), 36–43. https://doi.org/10.1016/j.neuroimage.2014.03.080

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

Cone, J., & Rand, D. G. (2014). Time pressure increases cooperation in competitively framed social dilemmas. *PLOS ONE*, *9*(12), e115756. https://doi.org/10.1371/journal.pone.0115756

Connor, R. C. (1995). The Benefits of Mutualism: A Conceptual Framework. *Biological Reviews*, *70*(3), 427–457. https://doi.org/10.1111/j.1469-185X.1995.tb01196.x

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314. https://doi.org/10.1037/0022-3514.83.6.1314

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*(4), 173–179. https://doi.org/10.1016/S1364-6613(03)00057-3

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281. https://doi.org/10.1016/S0899-8256(03)00119-2

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dawes, C. T., Loewen, P. J., Schreiber, D., Simmons, A. N., Flagan, T., McElreath, R., … Paulus, M. P. (2012). Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences*, *109*(17), 6479–6483. https://doi.org/10.1073/pnas.1118653109

de Botton, A. (2012). *Religion for Atheists: a non-believer's guide to the uses of religion*. New York, NY: Pantheon Books.

De Cremer, D., Van Knippenberg, D., Van Dijk, E., & Van Leeuwen, E. (2008). Cooperating if one's goals are collective-based: social identification effects in

social dilemmas as a function of goal transformation. *Journal of Applied Social Psychology*, *38*(6), 1562–1579. https://doi.org/10.1111/j.1559-1816.2008.00359.x

De Cremer, D., & Van Vugt, M. (1999). Social identification effects in social dilemmas: a transformation of motives. *European Journal of Social Psychology*, *29*(7), 871–893. https://doi.org/10.1002/(SICI)1099-0992

De Dreu, C. K. W., Balliet, D., Halevy, N., & Elliot, A.J. (2014). Parochial cooperation in humans: forms and functions of self-sacrifice in intergroup conflict. *Advances in Motivation Science.*, 1–47. https://doi.org/10.1016/bs.adms.2014.08.001

De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of Personality and Social Psychology*, *90*(6), 927–943. https://doi.org/10.1037/0022-3514.90.6.927

De Dreu, C. K. W., Dussel, D. B., & Velden, F. S. T. (2015). In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00572

De Dreu, C. K. W., Emans, B. J. M., Vliert, E., & Carnevale, P. J. (1994). Effects of gain-loss frames in negotiation: Loss aversion, mismatching, and frame adoption. *Organizational Behavior and Human Decision Processes*, *60*, 90–107. https://doi.org/10.1006/obhd.1994.1076

De Dreu, C. K. W., Emans, B. J., & Van de Vliert, E. (1992). Frames of reference and cooperative social decision-making. *European Journal of Social Psychology*, *22*(3), 297–302. https://doi.org/10.1002/ejsp.2420220307

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Kleef, G. A. V., Baas, M., … Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*(5984), 1408–1411. https://doi.org/10.1126/science.1189047

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences*, *108*(4), 1262–1266. https://doi.org/10.1073/pnas.1015316108

De Dreu, C. K. W., Gross, J., Méder, Z., Giffin, M., Prochazkova, E., Krikeb, J., & Columbus, S. (2016). In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences*, *113*(38), 10524–10529. https://doi.org/10.1073/pnas.1605115113

De Dreu, C. K. W., & McCusker, C. (1997). Gain–loss frames and cooperation in two-person social dilemmas: A transformational analysis. *Journal of Personality and Social Psychology*, *72*(5), 1093. https://doi.org/10.1037/0022-3514.72.5.1093

De Dreu, C. K. W., & Van Lange, P. A. M. (1995). The impact of social value orientations on negotiator cognition and behavior. *Personality and Social Psychology Bulletin*, *21*(11), 1178–1188. https://doi.org/10.1177/01461672952111006

de Heus, P., Hoogervorst, N., & Dijk, E. van. (2010). Framing prisoners and chickens: Valence effects in the prisoner's dilemma and the chicken game. *Journal of Experimental Social Psychology*, *46*(5), 736–742. https://doi.org/10.1016/j.jesp.2010.04.013

Deutsch, M. (1949). An experimental study of the effects of co-operation and competition upon group process. *Human Relations*, *2*(3), 199–231. https://doi.org/10.1177/001872674900200301

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. (2006). *The social psychology of prosocial behavior* (Vol. xiv). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Dunegan, K. J. (1993). Framing, cognitive modes, and image theory: Toward an understanding of a glass half full. *Journal of Applied Psychology*, *78*(3), 491–503. https://doi.org/10.1037/0021-9010.78.3.491

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(7), 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Eckel, C. C., & Grossman, P. J. (2003). Rebate versus matching: does how we subsidize charitable contributions matter? *Journal of Public Economics*, *87*(3–4), 681–701. https://doi.org/10.1016/S0047-2727(01)00094-9

Eckel, C. C., & Grossman, P. J. (2004). Giving to secular causes by the religious and nonreligious: An experimental test of the responsiveness of giving to subsidies. *Nonprofit and Voluntary Sector Quarterly*, *33*(2), 271–289. https://doi.org/10.1177/0899764004263423

Edgell, P., Gerteis, J., & Hartmann, D. (2006). Atheists as "other": Moral boundaries and cultural membership in American society. *American Sociological Review*, *71*(2), 211–234. https://doi.org/10.1177/000312240607100203

Ellemers, N., & Haslam, S. A. (2011). Social identity theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 379–398). London, England: Sage.

Ellemers, N., Spears, R., & Doosje, B. (1997). Sticking together or falling apart: In-group identification as a psychological determinant of group commitment versus individual mobility. *Journal of Personality and Social Psychology*, *72*(3), 617. https://doi.org/10.1037/0022-3514.72.3.617

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Ensari, N., & Miller, N. (2001). Decategorization and the reduction of bias in the cross categorization paradigm. *European Journal of Social Psychology*, *31*(2), 193–216. https://doi.org/10.1002/ejsp.42

Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology. General*, *145*(6), 772–787. https://doi.org/10.1037/xge0000165

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

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083. https://doi.org/10.1038/nature07155

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. In S.-C. Kolm & J. M. Ythier (Eds.), *Handbook of the economics of giving, altruism and reciprocity* (Vol. 2, pp. 615–691). Amsterdam, NE: Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S1574071406010086

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369. https://doi.org/10.1006/game.1994.1021

Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: Categorization, reciprocation, or fear? *Journal of Personality and Social Psychology*, *79*(1), 77–94. https://doi.org//10.1037/0022-3514.79.1.77

Gaertner, L., Iuzzini, J., Witt, M. G., & Oriña, M. M. (2006). Us without them: evidence for an intragroup origin of positive in-group regard. *Journal of Personality and Social Psychology*, *90*(3), 426–439. https://doi.org/10.1037/0022-3514.90.3.426

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics: Theory, Research, and Practice*, *4*(1), 98–114. https://doi.org/10.1037/1089-2699.4.1.98

Galen, L. W. (2012). Does religious belief promote prosociality? A critical examination. *Psychological Bulletin*, *138*(5), 876–906. https://doi.org/10.1037/a0028251

Gervais, W. M. (2011). Finding the faithless: perceived atheist prevalence reduces anti-atheist prejudice. *Personality and Social Psychology Bulletin*, *37*(4), 543–56. https://doi.org/10.1177/0146167211399583

Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology*, *101*(6), 1189–206. https://doi.org/10.1037/a0025882

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212–216. https://doi.org/10.1257/000282806777211658

Gonzalez, C., Dana, J., Koshino, H., & Just, M. (2005). The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology*, *26*(1), 1–20. https://doi.org/10.1016/j.joep.2004.08.004

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: the strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Güth, W., Ploner, M., & Regner, T. (2009). Determinants of in-group bias: Is group affiliation mediated by guilt-aversion? *Journal of Economic Psychology*, *30*(5), 814–827. https://doi.org/10.1016/j.joep.2009.07.001

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review*, *101*(4), 709–725. https://doi.org/10.1017/S0003055407070499

Hackel, L., Zaki, J., & Van Bavel, J. J. (2017). Social identity shapes social valuation: evidence from prosocial behavior and vicarious reward. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsx045

Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: a new game

paradigm. *Psychological Science*, *19*(4), 405–411. https://doi.org/10.1111/j.1467-9280.2008.02100.x

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making*, *25*(2), 188–195. https://doi.org/10.1002/bdm.726

Hall, D. L., Matz, D. C., & Wood, W. (2010). Why don't we practice what we preach? A meta-analytic review of religious racism. *Personality and Social Psychology Review*, *14*(1), 126–139. https://doi.org/10.1177/1088868309352179

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*(1), 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hanley, J. A., Negassa, A., Edwardes, M. D. deB, & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American Journal of Epidemiology*, *157*(4), 364–375. https://doi.org/10.1093/aje/kwf215

Harcourt, A. H., & de Waal, F. B. (Eds.). (1992). *Coalitions and alliances in humans and other animals*. Oxford, England: Oxford University Press. Retrieved from https://www.researchgate.net/profile/Ronald_Noe/publication/291900929_Alliance_formation_among_male_baboons_Shopping_for_profitable_partners/links/57bd972408ae691824302ab4/Alliance-formation-among-male-baboons-Shopping-for-profitable-partners.pdf

Hardin, G. (1968). The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science*, *162*(3859), 1243–1248. https://doi.org/10.1080/19390450903037302

Hardin, J. W. (2005). Generalized Estimating Equations (GEE). In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa250/abstract

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 175–204). San Diego, CA, US: Academic Press. https://doi.org/10.1016/B978-012208930-5/50007-6

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, *53*(1), 3–35. https://doi.org/10.1016/S0167-2681(03)00094-5

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

Henrich, N., & Henrich, J. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford: Oxford University Press.

Herek, G. M. (2000). The psychology of sexual prejudice. *Current Directions in Psychological Science*, *9*(1), 19–22. https://doi.org/10.1111/1467-8721.00051

Hertwig, R., & Ortmann, A. (2008). Deception in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, *71*(3), 222–227. https://doi.org/10.1177/019027250807100304

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

Hey, J. D. (1991). *Experiments in economics*. Oxford, England: Blackwell.

Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, *19*(3), 397–401. https://doi.org/10.1016/S0167-4870(98)00013-0

Hintze, A., Olson, R. S., Adami, C., & Hertwig, R. (2015). Risk sensitivity as an evolutionary adaptation. *Scientific Reports*, *5*, srep08242. https://doi.org/10.1038/srep08242

Hitchens, C. (2007). *God Is Not Great: How Religion Poisons Everything*. New York, NY: Twelve Books.

Hobbes, T. (1994). *Leviathan*. Hackett. (Original work published 1668)

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, *7*(3), 346–380. https://doi.org/10.1006/game.1994.1056

Hogg, M. A. (1993). Group cohesiveness: A critical review and some new directions. *European Review of Social Psychology*, *4*(1), 85–111. https://doi.org/10.1080/14792779343000031

Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology*, *11*(1), 223–255. https://doi.org/10.1080/14792772043000040

Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes* (Vol. xv). Florence, KY, US: Taylor & Frances/Routledge.

Hogg, M. A., & Abrams, D. (1993). Towards a single-process uncertainty-reduction model of social motivation in groups. In M. A. Hogg & D. Abrams (Eds.), *Group motivation: Social psychological perspectives* (pp. 173–190). Hertfordshire, HP2 7EZ, England: Harvester Wheatsheaf.

Holmes, J. G. (2002). Interpersonal expectations as the building blocks of social cognition: an interdependence theory perspective. *Personal Relationships*, *9*(1), 1–26. https://doi.org/10.1111/1475-6811.00001

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425. https://doi.org/10.1007/s10683-011-9273-9

Hout, M., & Fischer, C. S. (2002). Why more Americans have no religious preference: Politics and generations. *American Sociological Review*, 165–190.

Huber, S., & Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions*, *3*(3), 710–724. https://doi.org/10.3390/rel3030710

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: the role of prejudice and facial affect in race categorization. *Psychological Science*, *15*(5), 342–345. https://doi.org/10.1111/j.0956-7976.2004.00680.x

Insko, C. A., & Schopler, J. (1998). Differential distrust of groups and individuals. In C. Sedikides, J. Schopler, & C. Asko (Eds.), *Intergroup cognition and intergroup behavior* (pp. 75–107). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Jackson, J. W. (2008). Reactions to social dilemmas as a function of group identity, rational calculations, and social context. *Small Group Research*, *39*(6), 673–705. https://doi.org/10.1177/1046496408322761

James, W. (1985). *The varieties of religious experience: A study in human nature*. Cambridge, MA: Harvard University Press. (Original work published 1902)

Jin, N., & Yamagishi, T. (1997). Group heuristics in social dilemma. *Japanese Journal of Social Psychology*, *12*(3), 190–198.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Joireman, J. A., Van Lange, P. A. M., Kuhlman, D. M., Van Vugt, M., & Shelley, G. P. (1997). An interdependence analysis of commuting decisions. *European Journal of Social Psychology*, *27*(4), 441–463. https://doi.org/10.1002/(SICI)1099-0992(199707)27:4<441::AID-EJSP804>3.0.CO;2-S

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY, US: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2007). Frames and brains: elicitation and control of response tendencies. *Trends in Cognitive Sciences*, *11*(2), 45–46. https://doi.org/10.1016/j.tics.2006.11.007

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, *59*(4), S285–S300.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kant, I. (2002). *Groundwork for the Metaphysics of Morals*. New Haven, CT: Yale University Press. (Original work published 1797)

Karp, D., Jin, N., Yamagishi, T., & Shinotsuka, H. (1993). Raising the minimum in the minimal group paradigm. *The Japanese Journal of Experimental Social Psychology*, *32*(3), 231–240. https://doi.org/10.2130/jjesp.32.231

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & M, A. (2003). *An atlas of interpersonal situations* (Vol. xii). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511499845

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal Relations*. New York, NY: Wiley.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31–F36. https://doi.org/10.1111/j.1467-7687.2005.0434a.x

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). "Native" objects and collaborators: infants' object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development*, *13*(1), 67–81. https://doi.org/10.1080/15248372.2011.567200

Kiyonari, T., & Yamagishi, T. (2004). Ingroup cooperation and the social exchange heuristic. In D. V. Suleiman, D. V. Budescu, I. Fischer, & D. M. Messick (Eds.), *Contemporary Psychological Research on Social Dilemmas* (pp. 269–286). Cambridge, England: Cambridge University Press.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klor, E. F., & Shayo, M. (2010). Social identity and preferences over redistribution. *Journal of Public Economics*, *94*(3–4), 269–278. https://doi.org/10.1016/j.jpubeco.2009.12.003

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908–911. https://doi.org/10.1038/nature05631

Komorita, S. S., & Carnevale, P. (1992). Motivational arousal vs. decision framing in social dilemmas. In W. B. G. Liebrand, D. M. Messick, & H. A. M. Wilke (Eds.), *Social dilemmas: Theoretical issues and research findings* (pp. 209–223). Elmsford, NY, US: Pergamon Press.

Komorita, S. S., & Parks, C. D. (1995). Interpersonal relations: Mixed-motive interaction. *Annual Review of Psychology*, *46*(1), 183–207. https://doi.org/10.1146/annurev.ps.46.020195.001151

Kramer, R. M., & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology*, *46*(5), 1044–1057. https://doi.org/10.1037/0022-3514.46.5.1044

Krebs, D. (2008). Morality: an evolutionary account. *Perspectives on Psychological Science*, *3*(3), 149–172. https://doi.org/10.1111/j.1745-6924.2008.00072.x

Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55. https://doi.org/10.1006/obhd.1998.2781

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, *127*(2), 187–208. https://doi.org/10.1037/0033-2909.127.2.187

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology. General*, *145*(8), 1001–1016. https://doi.org/10.1037/xge0000179

Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of Personality and Social Psychology*, *84*(5), 932–943. https://doi.org/10.1037/0022-3514.84.5.932

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth, *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Leidner, B., Castano, E., Zaiser, E., & Giner-Sorolla, R. (2010). Ingroup glorification, moral disengagement, and justice in the context of collective violence. *Personality and Social Psychology Bulletin*, *36*(8), 1115–1129. https://doi.org/10.1177/0146167210376391

Leliveld, M. C., Beest, I. van, Dijk, E. van, & Tenbrunsel, A. E. (2009). Understanding the influence of outcome valence in bargaining: A study on fairness accessibility, norms, and behavior. *Journal of Experimental Social Psychology*, *45*(3), 505–514. https://doi.org/10.1016/j.jesp.2009.02.006

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Levine, J. M., & Moreland, R. L. (1994). Group socialization: Theory and research. *European Review of Social Psychology*, *5*(1), 305–336. https://doi.org/10.1080/14792779543000093

LeVine, R. A., & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior.* Oxford, England: John Wiley & Sons.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, *21*(2), 153–174.

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill.

Lewis, V. A., MacGregor, C. A., & Putnam, R. D. (2013). Religion, networks, and neighborliness: The impact of religious social networks on civic engagement. *Social Science Research*, *42*(2), 331–346. https://doi.org/10.1016/j.ssresearch.2012.09.011

Li, Y. J., Kenrick, D. T., Griskevicius, V., & Neuberg, S. L. (2012). Economic decision biases and fundamental motivations: how mating and self-protection alter loss aversion. *Journal of Personality and Social Psychology*, *102*(3), 550–561. https://doi.org/10.1037/a0025844

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*(1), 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482–493. https://doi.org/10.1086/519249

Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology*, *39*(5), 773–783. https://doi.org/10.1037/0022-3514.39.5.773

Ma, Y., Liu, Y., Rand, D. G., Heatherton, T. F., & Han, S. (2015). Opposing oxytocin effects on intergroup cooperative behavior in intuitive and reflective minds. *Neuropsychopharmacology*, *40*(10), 2379–2387. https://doi.org/10.1038/npp.2015.87

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The "Black Sheep Effect": Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*(1), 1–16. https://doi.org/10.1002/ejsp.2420180102

McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, *61*(2), 190–201. https://doi.org/10.1006/obhd.1995.1015

McDermott, R., Fowler, J. H., & Smirnov, O. (2008). On the Evolutionary Origin of Prospect Theory Preferences. *The Journal of Politics*, *70*(2), 335–350. https://doi.org/10.1017/S0022381608080341

McKay, R., Efferson, C., Whitehouse, H., & Fehr, E. (2011). Wrath of God: religious primes and punishment. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1713), 1858–63. https://doi.org/10.1098/rspb.2010.2125

McNeil, B. J., Pauker, S. G., Sox, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *The New England Journal of Medicine*, *306*(21), 1259–1262. https://doi.org/10.1056/NEJM198205273062103

Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For members only: Ingroup punishment of fairness norm violations in the ultimatum game. *Social Psychological and Personality Science*, *5*(6), 662–670. https://doi.org/10.1177/1948550614527115

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4, pp. 11–44). Beverly Hills, CA: Sage.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25. https://doi.org/10.1016/0022-1031(68)90046-2

Mifune, N., Hashimoto, H., & Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior*, *31*(2), 109–117. https://doi.org/10.1016/j.evolhumbehav.2009.09.004

Mill, J. S. (1863). *Utilitarianism*. London, England: Parker, Son, and Bourne.

Miller, S. L., Zielaskowski, K., & Plant, E. A. (2012). The basis of shooter biases: Beyond cultural stereotypes. *Personality and Social Psychology Bulletin*, *38*(10), 1358–1366. https://doi.org/10.1177/0146167212450516

Mischel, W. (1972). Direct versus indirect personality assessment: Evidence and implications. *Journal of Consulting and Clinical Psychology*, *38*(3), 319–324. https://doi.org/10.1037/h0032896

Monsma, S. V. (2007). Religion and philanthropic giving and volunteering: Building blocks for civic responsibility. *Interdisciplinary Journal of Research on Religion*, *3*(1), 1–28.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/10.1002/ejsp.2420220202

Mummendey, A., & Otten, S. (1998). Positive–negative asymmetry in social discrimination. *European Review of Social Psychology*, *9*(1), 107–143. https://doi.org/10.1080/14792779843000063

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin*, *26*(10), 1258–1270.

Mummendey, A., Simon, B., Dietze, C., Grünert, M., Haeger, G., Kessler, S., … Schäferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*(2), 125–144. https://doi.org/10.1016/0022-1031(92)90035-I

Mussweiler, T., & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proceedings of the National Academy of Sciences*, *110*(48), 19318–19323. https://doi.org/10.1073/pnas.1215443110

Noë, R., & Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology*, *35*(1), 1–11. https://doi.org/10.1007/BF00167053

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, *322*(5898), 58–62. https://doi.org/10.1126/science.1158757

Nowak, M. A., & Sigmund, K. (1998). The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology*, *194*(4), 561–574. https://doi.org/10.1006/jtbi.1998.0775

Ockenfels, A., & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior & Organization*, *108*, 453–462. https://doi.org/10.1016/j.jebo.2013.12.003

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Orbell, J., Goldman, M., Mulford, M., & Dawes, R. (1992). Religion, context, and constraint toward strangers. *Rationality and Society*, *4*(3), 291–307. https://doi.org/10.1177/1043463192004003004

Otten, S., & Mummendey, A. (1999). To our benefit or at your expense? Justice considerations in intergroup allocations of positive and negative resources. *Social Justice Research*, *12*(1), 19–38. https://doi.org/10.1023/A:1023274223181

Otten, S., Mummendey, A., & Blanz, M. (1996). Intergroup discrimination in positive and negative outcome allocations: Impact of stimulus valence, relative group status, and relative group size. *Personality and Social Psychology Bulletin*, *22*(6), 568–581. https://doi.org/10.1177/0146167296226003

Paciotti, B., Richerson, P. J., Baum, B., Lubell, M., Waring, T., McElreath, R., … Edsten, E. (2011). Are religious individuals more generous, trusting, and cooperative? An experimental test of the effect of religion on prosociality. *Research in Economic Anthropology*, *31*, 267–305. https://doi.org/10.1108/S0190-1281(2011)0000031014

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Pemberton, M. B., Insko, C. A., & Schopler, J. (1996). Memory for and experience of differential competitive behavior of individuals and groups. *Journal of Personality and Social Psychology*, *71*(5), 953. https://doi.org/10.1037/0022-3514.71.5.953

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annual Review of Psychology*, *56*(1), 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin*, *23*(2), 173–185. https://doi.org/10.1177/0146167297232006

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*(1), 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nature Communications*, *5*, 4939. https://doi.org/10.1038/ncomms5939

Peysakhovich, A., & Rand, D. G. (2015). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*. https://doi.org/10.1287/mnsc.2015.2168

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The Content of Our Cooperation, Not the Color of Our Skin: An Alliance Detection System Regulates Categorization by Coalition and Race, but Not Sex. *PLOS ONE*, *9*(2), e88534. https://doi.org/10.1371/journal.pone.0088534

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, *28*(1), 363–392. https://doi.org/10.1146/annurev.ps.28.020177.002051

Putnam, R. D., & Campbell, D. E. (2012). *American grace: How religion divides and unites us*. New York, NY: Simon and Schuster.

Rabbie, J. M., & Horwitz, M. (1969). Arousal of ingroup-outgroup bias by a chance win or loss. *Journal of Personality and Social Psychology*, *13*(3), 269–277. https://doi.org/10.1037/h0028284

Rabbie, J. M., & Horwitz, M. (1988). Categories versus groups as explanatory concepts in intergroup relations. *European Journal of Social Psychology*, *18*(2), 117–123. https://doi.org/10.1002/ejsp.2420180204

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology*, *19*(3), 171–202. https://doi.org/10.1002/ejsp.2420190302

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, *27*(9), 1192–1206. https://doi.org/10.1177/0956797616654455

Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General*, *145*(4), 389–396. https://doi.org/10.1037/xge0000154

Rand, D. G., & Epstein, Z. G. (2014). Risking Your Life without a Second Thought: Intuitive Decision-Making and Extreme Altruism. *PLOS ONE*, *9*(10), e109687. https://doi.org/10.1371/journal.pone.0109687

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. https://doi.org/10.1038/nature11467

Rand, D. G., Greene, J. D., & Nowak, M. A. (2013). Rand et al. reply. *Nature*, *498*(7452), E2–E3. https://doi.org/10.1038/nature12195

Rand, D. G., Newman, G. E., & Wurzbacher, O. M. (2015). Social context and the dynamics of cooperative choice. *Journal of Behavioral Decision Making*, *28*(2), 159–166. https://doi.org/10.1002/bdm.1837

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. https://doi.org/10.1038/ncomms4677

Rand, D. G., Pfeiffer, T., Dreber, A., Sheketoff, R. W., Wernerfelt, N. C., & Benkler, Y. (2009). Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of the National Academy of Sciences*, *106*(15), 6187–6191. https://doi.org/10.1073/pnas.0811552106

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Richerson, P. J., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., … Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *The Behavioral and Brain Sciences*, *39*, e30. https://doi.org/10.1017/S0140525X1400106X

Richerson, P. J., & Boyd, R. (2005). *Not By Genes Alone*. Chicago, IL: University of Chicago Press. Retrieved from http://www.press.uchicago.edu/ucp/books/book/chicago/N/bo3615170.html

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, *3*(4), 263–277. https://doi.org/10.1002/bdm.3960030404

Rockenbach, B., & Milinski, M. (2011). To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive.

Proceedings of the National Academy of Sciences, *108*(45), 18307–18312. https://doi.org/10.1073/pnas.1108996108

Ruffle, B. J., & Sosis, R. (2006). Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents. *Journal of Economic Behavior & Organization*, *60*(2), 147–163. https://doi.org/10.1016/j.jebo.2004.07.007

Rusbult, C. E., & Van Lange, P. A. M. (2008). Why we need interdependence theory. *Social and Personality Psychology Compass*, *2*(5), 2049–2070. https://doi.org/10.1111/j.1751-9004.2008.00147.x

Rusch, H. (2014). The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1794), 20141539. https://doi.org/10.1098/rspb.2014.1539

Scanlon, T. M. (1998). *What we owe to each other* (Vol. 66). Cambridge, MA: Belknap Press of Harvard University Press.

Schelling, T. C. (1960). The reciprocal fear of surprise attack. In *The Strategy of Conflict* (pp. 207–229). Cambridge, MA: Harvard University Press.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172–177. https://doi.org/10.1016/j.jesp.2016.09.009

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Semmann, D., Krambeck, H.-J., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, *56*(3), 248–252. https://doi.org/10.1007/s00265-004-0782-9

Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). Mturk character misrepresentation: assessment and solutions. *Journal of Consumer Research*, *44*(1), 211–230. https://doi.org/10.1093/jcr/ucx053

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. Boston, MA: Houghton Mifflin.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: the Robbers Cave experiment*. Norman: University of Oklahoma Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior*, *25*(6), 379–393. https://doi.org/10.1016/j.evolhumbehav.2004.08.001

Simpson, B. (2006). Social identity and cooperation in social dilemmas. *Rationality and Society*, *18*(4), 443–470. https://doi.org/10.1177/1043463106066381

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: An experimental study of fear-based aggression. *Journal of Experimental Social Psychology*, *49*(6), 1120–1123. https://doi.org/10.1016/j.jesp.2013.08.003

Sloan Wilson, D., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*(4), 585–608. https://doi.org/10.1017/S0140525X00036104

Sloan Wilson, D., Van Vugt, M., & O'Gorman, R. (2008). Multilevel selection theory and major evolutionary transitions: implications for psychological science. *Current Directions in Psychological Science*, *17*(1), 6–9. https://doi.org/10.1111/j.1467-8721.2008.00538.x

Sloan Wilson, D., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology*, *82*(4), 327–348. https://doi.org/10.1086/522809

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Smith, E. R., Coats, S., & Walling, D. (1999). Overlapping mental representations of self, in-group, and partner: further response time evidence and a connectionist model. *Personality and Social Psychology Bulletin*, *25*(7), 873–882. https://doi.org/10.1177/0146167299025007009

Smith, R. H., Powell, C. A. J., Combs, D. J. Y., & Schurtz, D. R. (2009). Exploring the when and why of schadenfreude. *Social and Personality Psychology Compass*, *3*(4), 530–546. https://doi.org/10.1111/j.1751-9004.2009.00181.x

Sober, E., & Sloan Wilson, D. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA, US: Harvard University Press.

Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior: the psychological functions of self-esteem and cultural worldviews. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 24, pp. 93–159). Academic Press. https://doi.org/10.1016/S0065-2601(08)60328-7

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665.

Stroebe, K., Lodewijkx, H. F. M., & Spears, R. (2005). Do unto others as they do unto you: Reciprocity and social identification as determinants of ingroup favoritism. *Personality and Social Psychology Bulletin*, *31*(6), 831–845. https://doi.org/10.1177/0146167204271659

Stürmer, S., & Snyder, M. (2009). Helping "us" versus "them." In S. Stürmer & M. Snyder (Eds.), *The Psychology of Prosocial Behavior* (pp. 33–58). Oxford, England: Wiley-Blackwell. https://doi.org/10.1002/9781444307948.ch2

Stürmer, S., Snyder, M., Kropp, A., & Siem, B. (2006). Empathy-motivated helping: The moderating role of group membership. *Personality and Social Psychology Bulletin*, *32*(7), 943–956. https://doi.org/10.1177/0146167206287363

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

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological Review*, *112*(1), 159–192. https://doi.org/10.1037/0033-295X.112.1.159

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–48). Monterey, CA: Brooks-Cole.

Tan, J. H. (2006). Religion and social preferences: An experimental study. *Economics Letters*, *90*(1), 60–67. https://doi.org/10.1016/j.econlet.2005.07.006

Tan, J. H., & Vogel, C. (2008). Religion and trust: An experimental study. *Journal of Economic Psychology*, *29*(6), 832–848. https://doi.org/10.1016/j.joep.2008.03.002

Ten Velden, F. S., Daughters, K., & De Dreu, C. K. W. (2017). Oxytocin promotes intuitive rather than deliberated cooperation with the in-group. *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2016.06.005

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups* (Vol. xiii). Oxford, England: John Wiley.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., … Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *498*(7452), E1–E2. https://doi.org/10.1038/nature12194

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human Morality & Sociality: Evolutionary & Comparative Perspectives* (pp. 191–234). New York, NY: Palgrave-Macmillan. Retrieved from http://www.cep.ucsb.edu/papers/groups%20in%20mind2010.pdf

Traulsen, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, *103*(29), 10952–10955. https://doi.org/10.1073/pnas.0602530103

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), 1089–1091. https://doi.org/10.1038/nature08785

Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.2307/2822435

Tropp, L. R., & Wright, S. C. (1999). Ingroup identification and relative deprivation: An examination across multiple social comparisons. *European Journal of Social Psychology*, *29*(5–6), 707–724. https://doi.org/10.1002/(SICI)1099-0992(199908/09)29:5/6<707::AID-EJSP968>3.0.CO;2-Y

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, *27*(5), 585–600. https://doi.org/10.1177/0146167201275007

Turner, J. C. (1982). Towards a cognitive redefinition of the social group. In H. Tajfel, *Social Identity and Intergroup Relations* (pp. 15–40). Cambridge, England: Cambridge University Press.

Turner, J. C., & Bourhis, R. Y. (1996). Social identity, interdependence, and the social group: a reply to Rabbie et al. In W. P. Robinson, *Social Groups and Identities: Developing the Legacy of Henri Tajfel*. Oxford, England: Butterworth-Heineman.

Turner, J. C., & Giles, H. (1981). *Intergroup behaviour*. Oxford, England: Blackwell.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Cambridge, MA, US: Basil Blackwell.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Van Beest, I., Van Dijk, E., De Dreu, C. K. W., & Wilke, H. A. (2005). Do-no-harm in coalition formation: Why losses inhibit exclusion and promote fairness cognitions. *Journal of Experimental Social Psychology*, *41*(6), 609–617. https://doi.org/10.1016/j.jesp.2005.01.002

Van Lange, P. A. M. (2012). A history of interdependence: Theory and research. In A. W. Kruglanski & W. Stroebe, *Handbook of the History of Social Psychology* (pp. 341–362). New York, NY: Psychology Press.

Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Vugt, M. V. (2007). From games to giving: social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology*, *29*(4), 375–384. https://doi.org/10.1080/01973530701665223

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125–141. https://doi.org/10.1016/j.obhdp.2012.11.003

Van Lange, P. A. M., & Liebrand, W. B. G. (1991). Social value orientation and intelligence: A test of the goal prescribes rationality principle. *European Journal of Social Psychology*, *21*(4), 273–292. https://doi.org/10.1002/ejsp.2420210402

Van Lange, P. A. M., Otten, W., De Bruin, E. M., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746. https://doi.org/10.1037/0022-3514.73.4.733

Van Lange, P. A. M., & Rusbult, C. E. (2011). Interdependence Theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 251–272). London, England: Sage.

Van Vugt, M., & Hart, C. M. (2004). Social identity as social glue: the origins of group loyalty. *Journal of Personality and Social Psychology*, *86*(4), 585–598. https://doi.org/10.1037/0022-3514.86.4.585

Van Vugt, M., & Park, J. H. (2009). The Tribal Instinct Hypothesis. In S. Stürmer & M. Snyder (Eds.), *The Psychology of Prosocial Behavior* (pp. 11–32). Oxford, England: Wiley-Blackwell.

Verkoeijen, P. P. J. L., & Bouwmeester, S. (2014). Does intuition cause cooperation? *PLOS ONE*, *9*(5), e96654. https://doi.org/10.1371/journal.pone.0096654

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (Vol. xviii). Princeton, NJ, US: Princeton University Press.

Waal, F. B. M. de, Leimgruber, K., & Greenberg, A. R. (2008). Giving is self-rewarding for monkeys. *Proceedings of the National Academy of Sciences*, *105*(36), 13685–13689. https://doi.org/10.1073/pnas.0807060105

Wåhlberg, A. E. af, & Poom, L. (2015). An empirical test of nonresponse bias in internet surveys. *Basic and Applied Social Psychology*, *37*(6), 336–347. https://doi.org/10.1080/01973533.2015.1111212

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. https://doi.org/10.1126/science.1121448

West, S. G., & Brown, T. J. (1975). Physical attractiveness, the severity of the emergency and helping: A field experiment and interpersonal simulation. *Journal of Experimental Social Psychology*, *11*(6), 531–538. https://doi.org/10.1016/0022-1031(75)90004-9

Whitney, P., Rinehart, C. A., & Hinson, J. M. (2008). Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic Bulletin & Review*, *15*(6), 1179–1184. https://doi.org/10.3758/PBR.15.6.1179

Whitt, S., & Wilson, R. K. (2007). The dictator game, fairness and ethnicity in postwar Bosnia. *American Journal of Political Science*, *51*(3), 655–668. https://doi.org/10.1111/j.1540-5907.2007.00273.x

Wilder, D. A. (1981). Perceiving persons as a group: Categorization and intergroup relations. In D. L. Hamilton, *Cognitive processes in stereotyping and intergroup behavior* (pp. 213–257). Hillsdale, NJ: Erlbaum.

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., & Schopler, J. (2003). Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin*, *129*(5), 698. https://doi.org/10.1037/0033-2909.129.5.698

Wit, A. P., & Kerr, N. L. (2002). "Me versus just us versus us all" categorization and cooperation in nested social dilemmas. *Journal of Personality and Social Psychology*, *83*(3), 616–637. https://doi.org/10.1037/0022-3514.83.3.616

Xygalatas, D. (2013). Effects of religious setting on cooperative behavior: a case study from Mauritius. *Religion, Brain & Behavior*, *3*(2), 91–102.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in Group Processes*, *16*(1), 161–197.

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, *63*(2), 116–132. https://doi.org/10.2307/2695887

Yamagishi, T., & Mifune, N. (2008). Does shared group membership promote altruism? Fear, greed, and reputation. *Rationality and Society*, *20*(1), 5–30. https://doi.org/10.1177/1043463107085442

Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: in-group love or out-group hate? *Evolution and Human Behavior*, *30*(4), 229–237. https://doi.org/10.1016/j.evolhumbehav.2009.02.004

Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of Experimental Psychology: General*, *145*(11), 1448–1459. https://doi.org/10.1037/xge0000190

Zahavi, A., & Zahavi, A. (1999). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford University Press.

Zaki, J., López, G., & Mitchell, J. P. (2014). Activity in ventromedial prefrontal cortex co-varies with revealed social preferences: evidence for person-invariant value. *Social Cognitive and Affective Neuroscience*, *9*(4), 464–469. https://doi.org/10.1093/scan/nst005

Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences*, *108*(49), 19761–19766. https://doi.org/10.1073/pnas.1112324108

Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science*, *22*(6), 466–470. https://doi.org/10.1177/0963721413492764

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions.

Journal of Personality and Social Psychology, *111*(4), 493–504. https://doi.org/10.1037/pspa0000056

Appendix

Study 1 (Dictator Game)

Analyses with only Atheists and Christians (*N* = 296)

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 1.73$, $t(293) = 4.94$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = 0.02$, $t(293) = 0.05$, $p = .96$
3. Interaction Effect: $B = 0.67$, $t(292) = 1.89$, $p = .06$
    a. Simple Effects for Atheist Participants: $B = -0.83$, $p = .19$
    b. Simple Effects for Religious Participants: $B = 0.86$, $p = .17$

Analysis with CRS controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $B = 1.94$, $t(280) = 5.04$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.09$, $t(280) = -0.21$, $p = .84$
3. Interaction Effect: $B = 0.58$, $t(279) = 1.61$, $p = .11$

Analysis using binary identification as Atheist or Religious

1. Main Effect of Participant Religiosity: $F(1,292) = 12.39$, $p < .001$
2. Main Effect of Recipient Religiosity: $F(1,292) = 0.02$, $p = .89$
3. Interaction Effect: $F(1,292) = 6.28$, $p = .01$
    a. Simple Effects for Atheist Participants: $t(128.41) = 2.19$, $p = .03$
    b. Simple Effects for Religious Participants: $t(163) = -1.58$, $p = .12$
    c. [*NB*: This effect is such that atheist participants transfer more to a partner who shares their atheist identification ($M = 8.42$) than a religious partner ($M = 6.06$), while religious participants don't show a difference]

Analysis using binary identification, controlling for gender, ethnicity, education, and political ideology

1. Main Effect of Participant Religiosity: $F(1,287) = 8.26$, $p = .004$
2. Main Effect of Recipient Religiosity: $F(1,287) = 0.06$, $p = .81$
3. Interaction Effect: $F(1,287) = 6.28$, $p = .01$

Analyses for Centrality of Religion Scale (CRS) Components Separately

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 2.02$, $t(293) = 5.49$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.13$, $t(293) = 0.29$, $p = .77$
3. Interaction Effect: $B = 0.47$, $t(293) = 1.25$, $p = .21$

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 1.12$, $t(293) = 4.34$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.04$, $t(293) = 0.10$, $p = .92$
3. Interaction Effect: $B = 0.57$, $t(292) = 2.23$, $p = .03$
    a. Simple Effects for Atheist Participants: $B = -0.19$, $p = .13$
    b. Simple Effects for Religious Participants: $B = 0.21$, $p = .10$
4. This relationship between religious ideology (believing that God exists) and transfer amounts was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.12$) not including zero (0.05, 0.20).

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 1.13$, $t(293) = 3.33$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.05$, $t(293) = -0.11$, $p = .92$
3. Interaction Effect: $B = 0.45$, $t(292) = 1.27$, $p = .21$
4. This relationship between public worship (attending religious services) and transfer amounts was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.22$) not including zero (0.12, 0.34).

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 1.15$, $t(293) = 4.03$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = 0.02$, $t(293) = 0.05$, $p = .96$
3. Interaction Effect: $B = 0.64$, $t(292) = 2.24$, $p = .03$
    a. Simple Effects for Atheist Participants: $B = -0.19$, $p = .13$
    b. Simple Effects for Religious Participants: $B = 0.20$, $p = .11$
4. This relationship between private worship (praying) and transfer amounts was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.15$) not including zero (0.07, 0.25).

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Recipient Religiosity: $B = 1.40$, $t(293) = 4.50$, $p < .001$

2. Main Effect of Participant Religiosity: $B = 0.00$, $t(293) = 0.00$, $p = .10$
3. Interaction Effect: $B = 0.47$, $t(292) = 1.51$, $p = .13$
4. This relationship between religious experience and transfer amounts was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.15$) not including zero (0.06, 0.25).

Analyses with only Christians ($N = 165$)

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 1.97$, $t(162) = 2.54$, $p = .01$
2. Main Effect of Recipient Religiosity: $B = 0.86$, $t(162) = 1.30$, $p = .19$
3. Interaction Effect: $B = 0.01$, $t(161) = 0.01$, $p = .10$

Analyses for Centrality of Religion Scale (CRS) Components Separately

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 2.29$, $t(162) = 3.49$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = 0.87$, $t(162) = 1.35$, $p = .18$
3. Interaction Effect: $B = 0.08$, $t(161) = 0.12$, $p = .91$

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 1.45$, $t(162) = 1.89$, $p = .06$
2. Main Effect of Recipient Religiosity: $B = 0.97$, $t(162) = 1.46$, $p = .15$
3. Interaction Effect: $B = -0.08$, $t(161) = -0.11$, $p = .92$

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 0.56$, $t(162) = 1.03$, $p = .30$
2. Main Effect of Recipient Religiosity: $B = 0.97$, $t(162) = 1.45$, $p = .15$
3. Interaction Effect: $B = 0.01$, $t(161) = 0.02$, $p = .99$

Analyses using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 0.86$, $t(162) = 1.56$, $p = .12$
2. Main Effect of Recipient Religiosity: $B = 0.97$, $t(162) = 1.47$, $p = .15$
3. Interaction Effect: $B = 0.23$, $t(161) = 0.42$, $p = .68$

Analyses using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: $B = 1.24$, $t(162) = 2.18$, $p = .03$

2. Main Effect of Recipient Religiosity: $B = 0.91$, $t(162) = 1.38$, $p = .17$
3. Interaction Effect: $B = -0.25$, $t(161) = -0.45$, $p = .66$

Analyses with all Religious Affiliations (N = 392)

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 1.64$, $t(389) = 4.97$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.16$, $t(389) = -0.41$, $p = .68$
3. Interaction Effect: $B = 0.64$, $t(388) = 1.93$, $p = .05$
    a. Simple Effects for Atheist Participants: $B = -0.91$, $p = .10$
    b. Simple Effects for Religious Participants: $B = 0.59$, $p = .28$

Analysis with CRS controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $B = 1.79$, $t(375) = 5.00$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.31$, $t(375) = -0.77$, $p = .045$
3. Interaction Effect: $B = 0.67$, $t(374) = 1.99$, $p = .05$
    a. Simple Effects for Atheist Participants: $B = -1.09$, $p = .05$
    b. Simple Effects for Religious Participants: $B = 0.48$, $p = .39$

Analyses using binary identification as Atheist or Religious

1. Main Effect of Participant Religiosity: $F(1,388) = 10.12$, $p = .002$
2. Main Effect of Recipient Religiosity: $F(1,388) = 0.92$, $p = .34$
3. Interaction Effect: $F(1,388) = 3.54$, $p = .06$
    a. Simple Effects for Atheist Participants: $t(128.41) = 2.19$, $p = .03$
    b. Simple Effects for Religious Participants: $t(259) = -0.74$, $p = .46$
    c. [*NB*: This effect is such that atheist participants transfer more to a partner who shares their atheist identification ($M = 8.42$) than a religious partner ($M = 6.06$), while religious participants don't show a difference]

Analyses using binary identification controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $F(1,383) = 8.03$, $p = .005$
2. Main Effect of Recipient Religiosity: $F(1,383) = 1.26$, $p = .26$
3. Interaction Effect: $F(1,388) = 3.79$, $p = .05$

Analyses for Centrality of Religion Scale (CRS) Components Separately.

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 1.75$, $t(389) = 5.20$, $p < .001$

2. Main Effect of Recipient Religiosity: B = -0.11, t(389) = -0.29, p = .77
3. Interaction Effect: B = 0.69, t(388) = 2.01, p = .05
    a. Simple Effects for Atheist Participants: B = -0.90, p = .10
    b. Simple Effects for Religious Participants: B = 0.66, p = .22

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: B = 1.03, t(389) = 4.26, p < .001
2. Main Effect of Recipient Religiosity: B = -0.11, t(389) = -0.28, p = .78
3. Interaction Effect: B = 0.59, t(388) = 2.44, p = .02
    a. Simple Effects for Atheist Participants: B = -1.59, p = .05
    b. Simple Effects for Religious Participants: B = 0.83, p = .13
4. This relationship between believing in God and transfer amounts in the DG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.10) not including zero (0.04, 0.16).

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: B = 1.02, t(389) = 3.14, p = .002
2. Main Effect of Recipient Religiosity: B = -0.24, t(389) = -0.62, p = .54
3. Interaction Effect: B = 0.28, t(388) = 0.86, p = .39
4. This relationship between public worship and transfer amounts in the DG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.18) not including zero (0.08, 0.28).

Analyses using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: B = 1.00, t(389) = 3.77, p < .001
2. Main Effect of Recipient Religiosity: B = -0.17, t(389) = -0.43, p = .67
3. Interaction Effect: B = 0.61, t(388) = 2.29, p = .02
    a. Simple Effects for Atheist Participants: B = -1.00, p = .06
    b. Simple Effects for Religious Participants: B = 0.73, p = .19
4. This relationship between frequency of praying and transfer amounts in the DG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.13) not including zero (0.06, 0.21).

Analyses using the fifth component of the CRS (Religious experience)

1. Main Effect of Recipient Religiosity: B = 1.43, t(389) = 5.00, p < .001

2. Main Effect of Participant Religiosity: B = -0.18, t(389) = -0.46, p = .65
3. Interaction Effect: B = 0.39, t(388) = 1.37, p = .17
4. This relationship between religious experience and transfer amounts in the DG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.11) not including zero (0.04, 0.19).

Study 2 (Trust Game)

Only Atheists and Christians (N = 296)

Transfer Amounts

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: B = 1.46, t(269) = 2.62, p =.009
2. Main Effect of Recipient Religiosity: B = -0.43, t(269) = -0.61, p =.55
3. Interaction Effect: B = 0.39, t(268) = 0.70, p =.48

Analysis with CRS controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: B = 1.22, t(257) = 1.97, p = .05
2. Main Effect of Recipient Religiosity: B = -0.39, t(257) = -0.39, p = .70
3. Interaction Effect: B = 0.20, t(256) = 0.35, p = .73

Analyses using binary identification as Atheist or Religious

1. Main Effect of Participant Religiosity: F(1,268) = 2.01, p = .16
2. Main Effect of Recipient Religiosity: F(1,268) = 0.25, p = .62
3. Interaction Effect: F(1,268) = 0.35, p = .56

Analyses using binary identification, controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: F(1,263) = 0.10, p = .76
2. Main Effect of Recipient Religiosity: F(1,263) = 0.08, p = .78
3. Interaction Effect: F(1,268) = 0.24, p = .62

Analyses for Centrality of Religion Scale (CRS) Components Separately

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 1.86$, $t(269) = 3.20$, $p = .002$
2. Main Effect of Recipient Religiosity: $B = -0.45$, $t(269) = -0.63$, $p = .53$
3. Interaction Effect: $B = 0.32$, $t(268) = 0.54$, $p = .59$

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 0.95$, $t(269) = 2.32$, $p = .02$
2. Main Effect of Recipient Religiosity: $B = -0.41$, $t(269) = -0.57$, $p = .57$
3. Interaction Effect: $B = 0.32$, $t(268) = 0.79$, $p = .43$
4. This relationship between religious ideology (believing that God exists) and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.61$) not including zero (0.07, 1.22).

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: B = 0.95, t(269) = 1.73, p = .09
2. Main Effect of Recipient Religiosity: B = -0.30, t(269) = -0.42, p = .68
3. Interaction Effect: B = 0.47, t(268) = 0.85, p = .39
4. This relationship between public worship (attending religious services) and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 1.15) not including zero (0.24, 2.09).

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: B = 0.96, t(269) = 2.09, p = .04
2. Main Effect of Recipient Religiosity: B = -0.38, t(269) = -0.53, p = .60
3. Interaction Effect: B = 0.41, t(268) = 0.88, p = .38
4. This relationship between private worship (praying) and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.85) not including zero (0.12, 1.69).

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: B = 1.21, t(269) = 2.39, p = .02
2. Main Effect of Recipient Religiosity: B = -0.48, t(269) = -0.65, p = .51
3. Interaction Effect: B = 0.08, t(268) = 0.17, p = .87
4. This relationship between religious experience and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95%

bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample (B = 0.78) not including zero (0.09, 1.58).

Predicted Returns

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: B = 3.52, t(269) = 3.03, p =.003
2. Main Effect of Recipient Religiosity: B = -1.19, t(269) = -0.79, p =.43
3. Interaction Effect: B = 2.27, t(268)= 1.96, p=.05
    a. Simple Effects for Atheist Participants: B = -4.12, p = .05,
    b. Simple Effects for Religious Participants: B = 1.70, p = .42

Analysis with CRS, controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: B = 3.00, t(257) = 2.30, p = .02
2. Main Effect of Recipient Religiosity: B = -0.87, t(257) = -0.57, p = .57
3. Interaction Effect: B = 2.09, t(256) = 1.72, p = .09
    a. Simple Effects for Atheist Participants: B = -3.53, p = .10,
    b. Simple Effects for Religious Participants: B = 1.76, p = .42

Analysis using binary identification as Atheist or Religious

1. Main Effect of Participant Religiosity: F(1,268) = 4.65, p = .03
2. Main Effect of Recipient Religiosity: F(1,268) = 0.54, p = .46
3. Interaction Effect: F(1,268) = 1.43, p = .23

Analysis using binary identification, controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: F(1,263) = 1.05, p = .31
2. Main Effect of Recipient Religiosity: F(1,263) = 0.28, p = .59
3. Interaction Effect: F(1,268) = 1.19, p = .28

Analyses for Centrality of Religion Scale (CRS) Components Separately.

Analysis using the first component of the CRS (Intellect: think about religious issues)

1. Main Effect of Participant Religiosity: B = 3.36, t(269) = 2.74, p = .007
2. Main Effect of Recipient Religiosity: B = -1.11, t(269) = -0.74, p = .46
3. Interaction Effect: B = 2.49, t(269) = 2.04, p = .04
    a. Simple Effects for Atheist Participants: B = -4.19, p = .05
    b. Simple Effects for Religious Participants: B = 1.89, p = .37

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: B = 2.43, $t(269)$ = 2.83, p = .005
2. Main Effect of Recipient Religiosity: B = -1.14, $t(269)$ = -0.76, p = .45
3. Interaction Effect: B = 1.57, $t(268)$ = 1.83, p = .07
    a. Simple Effects for Atheist Participants: B = -3.91, p = .07
    b. Simple Effects for Religious Participants: B = 1.55, p = .46

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: B = 2.35, $t(269)$ = 2.04, p = .04
2. Main Effect of Recipient Religiosity: B = -0.86, $t(269)$ = -0.57, p = .57
3. Interaction Effect: B = 2.08, $t(269)$ = 1.81, p = .07
    a. Simple Effects for Atheist Participants: B = -3.26, p = .10
    b. Simple Effects for Religious Participants: B = 1.84, p = .39

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: B = 2.49, $t(269)$ = 2.59, p = .01
2. Main Effect of Recipient Religiosity: B = -1.08, $t(269)$ = -0.72, p = .47
3. Interaction Effect: B = 1.48, $t(268)$ = 1.54, p = .13

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: B = 3.30, $t(269)$ = 3.14, p = .002
2. Main Effect of Recipient Religiosity: B = -1.36, $t(269)$ = -0.91, p = .36
3. Interaction Effect: B = 1.75, $t(268)$ = 1.67, p = .10

Only Christian Participants (N = 139)

Transfer Amounts

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: B = 2.67, $t(136)$ = 2.38, p =.02
2. Main Effect of Recipient Religiosity: B = 0.86, $t(136)$ = 1.30, p =.19
3. Interaction Effect: B = -0.04, $t(135)$ = 0.44 , p =.66

Analyses for Centrality of Religion Scale (CRS) Components Separately

Analysis using the first component of the CRS (Intellect: think about religious issues)

1. Main Effect of Participant Religiosity: B = 3.30, $t(136)$ = 3.49, $p < .001$
2. Main Effect of Recipient Religiosity: B = -0.21, $t(136)$ = -0.23, p = .82
3. Interaction Effect: B = 0.52, $t(135)$ = 0.54, p = .59

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 2.30$, $t(136) = 2.20$, $p = .03$
2. Main Effect of Recipient Religiosity: $B = 0.10$, $t(136) = 0.10$, $p = .92$
3. Interaction Effect: $B = 0.54$, $t(135) = 0.52$, $p = .61$

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 0.52$, $t(136) = 0.64$, $p = .53$
2. Main Effect of Recipient Religiosity: $B = 0.07$, $t(136) = .07$, $p = .94$
3. Interaction Effect: $B = 0.96$, $t(135) = 1.19$, $p = .24$

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 1.10$, $t(136) = 1.27$, $p = .21$
2. Main Effect of Recipient Religiosity: $B = 0.03$, $t(136) = 0.03$, $p = .98$
3. Interaction Effect: $B = 0.58$, $t(135) = 0.67$, $p = .51$

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: $B = 1.80$, $t(136) = 2.16$, $p = .03$
2. Main Effect of Recipient Religiosity: $B = -0.15$, $t(136) = -0.16$, $p = .87$
3. Interaction Effect: $B = -0.73$, $t(135) = -0.86$, $p = .39$

Predicted Returns

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 4.59$, $t(136) = 1.85$, $p = .07$
2. Main Effect of Recipient Religiosity: $B = 0.52$, $t(136) = 0.24$, $p = .81$
3. Interaction Effect: $B = 3.27$, $t(135) = 1.32$, $p = .19$

Analyses for Centrality of Religion Scale (CRS) Components Separately

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 6.39$, $t(136) = 3.04$, $p = .003$
2. Main Effect of Recipient Religiosity: $B = 0.16$, $t(136) = 0.08$, $p = .94$
3. Interaction Effect: $B = 3.51$, $t(135) = 1.68$, $p = .10$

Analyses using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 2.86$, $t(136) = 1.23$, $p = .22$
2. Main Effect of Recipient Religiosity: $B = 0.74$, $t(136) = 0.35$, $p = .73$

3. Interaction Effect: $B = 3.05$, $t(135) = 1.31$, $p = .19$

Analyses using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 0.45$, $t(136) = 0.25$, $p = .80$
2. Main Effect of Recipient Religiosity: $B = 0.70$, $t(136) = .33$, $p = .75$
3. Interaction Effect: $B = 2.80$, $t(135) = 1.58$, $p = .12$

Analyses using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 1.52$, $t(136) = 0.80$, $p = .42$
2. Main Effect of Recipient Religiosity: $B = 0.65$, $t(136) = 0.30$, $p = .77$
3. Interaction Effect: $B = 1.24$, $t(135) = 0.65$, $p = .52$

Analyses using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: $B = 4.09$, $t(136) = 2.24$, $p = .03$
2. Main Effect of Recipient Religiosity: $B = 0.21$, $t(136) = 0.10$, $p = .92$
3. Interaction Effect: $B = 0.53$, $t(135) = 0.28$, $p = .78$

Analyses with All Religious Affiliations ($N = 339$).

Transfer Amounts

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 1.24$, $t(336) = 2.41$, $p = .02$
2. Main Effect of Recipient Religiosity: $B = -0.36$, $t(336) = -0.58$, $p = .57$
3. Interaction Effect: $B = 0.39$, $t(335) = 0.75$, $p = .45$

Analysis with CRS controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $B = 1.11$, $t(322) = 1.97$, $p = .05$
2. Main Effect of Recipient Religiosity: $B = -0.27$, $t(322) = -0.43$, $p = .67$
3. Interaction Effect: $B = 0.19$, $t(321) = 0.36$, $p = .72$

Analysis using Binary Identification as Atheist or Religious

1. Main Effect of Participant Religiosity: $F(1,335) = 3.50$, $p = .06$
2. Main Effect of Recipient Religiosity: $F(1,335) = 0.48$, $p = .49$
3. Interaction Effect: $F(1,335) = 0.27$, $p = .60$

Analysis using binary identification controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $F(1,330) = 0.80$, $p = .37$
2. Main Effect of Recipient Religiosity: $F(1,330) = 0.25$, $p = .62$
3. Interaction Effect: $F(1,330) = 0.18$, $p = .68$

Analyses for Centrality of Religion Scale (CRS) Components Separately.

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 1.76$, $t(336) = 3.33$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -0.33$, $t(336) = -0.54$, $p = .59$
3. Interaction Effect: $B = 0.16$, $t(335) = 0.29$, $p = .77$

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 0.79$, $t(336) = 2.10$, $p = .04$
2. Main Effect of Recipient Religiosity: $B = -0.36$, $t(336) = -0.56$, $p = .57$
3. Interaction Effect: $B = 0.27$, $t(335) = 0.73$, $p = .47$
4. This relationship between believing in God and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.12$) not including zero (0.03, 0.22).

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 0.66$, $t(336) = 1.30$, $p = .20$
2. Main Effect of Recipient Religiosity: $B = -0.28$, $t(336) = -0.44$, $p = .66$
3. Interaction Effect: $B = 0.42$, $t(335) = 0.81$, $p = .42$
4. This relationship between public worship and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.24$) not including zero (0.09, 0.42).

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 0.78$, $t(336) = 1.87$, $p = .06$
2. Main Effect of Recipient Religiosity: $B = -0.33$, $t(336) = -0.52$, $p = .61$
3. Interaction Effect: $B = 0.43$, $t(335) = 1.02$, $p = .31$
4. This relationship between frequency of praying and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.17$) not including zero (0.04, 0.32).

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: $B = 0.94$, $t(336) = 2.04$, $p = .04$
2. Main Effect of Recipient Religiosity: $B = -0.41$, $t(336) = -0.65$, $p = .52$
3. Interaction Effect: $B = 0.19$, $t(335) = 0.42$, $p = .68$
4. This relationship between religious experience and transfer amounts in the TG was fully mediated by thinking about religious issues, with the 95% bias-corrected bootstrap confidence interval based on a 10,000 bootstrap sample ($B = 0.16$) not including zero (0.04, 0.30).

Predicted Returns in the TG

Analysis with Continuous Measure of Religiosity (CRS)

1. Main Effect of Participant Religiosity: $B = 3.32$, $t(336) = 3.02$, $p = .003$
2. Main Effect of Recipient Religiosity: $B = -1.11$, $t(336) = -0.83$, $p = .41$
3. Interaction Effect: $B = 2.18$, $t(335) = 1.20$, $p = .05$
    a. Simple Effects for Atheist Participants: $B = -3.78$, $p = .05$
    b. Simple Effects for Religious Participants: $B = 1.54$, $p = .42$

Analysis with CRS, controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $B = 2.79$, $t(322) = 2.30$, $p = .02$
2. Main Effect of Recipient Religiosity: $B = -0.70$, $t(322) = -0.50$, $p = .62$
3. Interaction Effect: $B = 1.83$, $t(321) = 1.61$, $p = .11$

Analysis using binary identification as Atheist or Religious

1. Main Effect of Participant Religiosity: $F(1,335) = 7.37$, $p = .007$
2. Main Effect of Recipient Religiosity: $F(1,335) = 1.17$, $p = .28$
3. Interaction Effect: $F(1,335) = 1.03$, $p = .31$

Analysis using binary identification, controlling for gender, ethnicity, education, and political ideology:

1. Main Effect of Participant Religiosity: $F(1,330) = 2.80$, $p = .10$
2. Main Effect of Recipient Religiosity: $F(1,330) = 0.55$, $p = .46$
3. Interaction Effect: $F(1,330) = 0.90$, $p = .34$

Analyses for Centrality of Religion Scale (CRS) Components Separately.

Analysis using the first component of the CRS (Intellect: thinking about religious issues)

1. Main Effect of Participant Religiosity: $B = 3.28$, $t(336) = 2.89$, $p = .004$
2. Main Effect of Recipient Religiosity: $B = -0.98$, $t(336) = -0.73$, $p = .46$

3. Interaction Effect: $B = 2.15$, $t(335) = 1.90$, $p = .06$
    a. Simple Effects for Atheist Participants: $B = -1.15$, $p = .06$
    b. Simple Effects for Religious Participants: $B = 1.21$, $p = .41$

Analysis using the second component of the CRS (Ideology: believing that God exists)

1. Main Effect of Participant Religiosity: $B = 2.16$, $t(336) = 2.70$, $p = .007$
2. Main Effect of Recipient Religiosity: $B = -1.10$, $t(336) = -0.81$, $p = .42$
3. Interaction Effect: $B = 1.35$, $t(335) = 1.68$, $p = .09$
    a. Simple Effects for Atheist Participants: $B = -3.37$, $p = .08$
    b. Simple Effects for Religious Participants: $B = 1.14$, $p = .55$

Analysis using the third component of the CRS (Public worship: attending religious services)

1. Main Effect of Participant Religiosity: $B = 1.89$, $t(336) = 1.74$, $p = .08$
2. Main Effect of Recipient Religiosity: $B = -0.89$, $t(336) = -0.66$, $p = .51$
3. Interaction Effect: $B = 2.11$, $t(335) = 1.94$, $p = .05$
    a. Simple Effects for Atheist Participants: $B = -3.12$, $p = .08$
    b. Simple Effects for Religious Participants: $B = 1.74$, $p = .36$

Analysis using the fourth component of the CRS (Private worship: praying)

1. Main Effect of Participant Religiosity: $B = 2.09$, $t(336) = 2.34$, $p = .02$
2. Main Effect of Recipient Religiosity: $B = -1.01$, $t(336) = -0.75$, $p = .45$
3. Interaction Effect: $B = 1.07$, $t(335) = 1.20$, $p = .23$

Analysis using the fifth component of the CRS (Religious experience)

1. Main Effect of Participant Religiosity: $B = 3.40$, $t(336) = 3.49$, $p < .001$
2. Main Effect of Recipient Religiosity: $B = -1.37$, $t(336) = -1.02$, $p = .31$
3. Interaction Effect: $B = 2.01$, $t(335) = 2.06$, $p = .04$
    a. Simple Effects for Atheist Participants: $B = -4.00$, $p = .03$
    b. Simple Effects for Religious Participants: $B = 1.39$, $p = .46$