

**Beyond the Dyad: The Role of Groups and Third-Parties
in the Trajectory of Violence**

Submitted by Richard Philpot to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Psychology
in June 2017

This thesis is available for Library use on the understanding that it is copyright material
and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that
no material has previously been submitted and approved for the award of a degree by this or
any other University.

Signature:

Acknowledgements

This thesis is dedicated to those who have provided invaluable support, guidance, and motivation towards its development over the past three (and a one-third) years.

However, prior to first panting up the Washington Singer mound (perhaps indicative of the incremental pressures to come), I had already a lot to be thankful for. I am indebted to my parents, Kevin and Linda, for their self-sacrifice to provide my sisters and I with the opportunities that they themselves truly deserve. I am sincerely grateful for the unconditional love of my two beloved children, Yena and Teeuw, both of whom motivate me in every productive thing I do. I am eternally indebted to my brilliant wife Leah, not only for her unwavering support throughout my postgraduate pursuits but also for her incalculable sacrifices. To be willing to move away from her family and invest all of our post-marital savings, just so I can wake up every day and enjoy my work—I am truly blessed. I also thank my parents-in-law, Gilbert and Sunny, for always believing in me, and for always worrying about my weight and wellbeing. I would like to express my gratitude to my previous MSc supervisor Steve Reicher, who convinced me not to settle for any postgraduate job but to follow my academic interests.

Here in Exeter, I wholeheartedly thank my three supervisors Mark, Miriam, and Richard—without whom none of this thesis would exist. I sincerely thank Mark Levine, not only for his remarkable academic support but also for his contagious relaxant properties. I would also like to thank Mark for giving me confidence as a developing researcher, and for providing numerous opportunities to meet interesting new people. Similar, profound gratitude goes to Miriam Koschate-Reis, for her gentle rigorousness and for always encouraging me to be critical of my own work and that of others. I express sincere thanks to Richard Everson, who not only survived three years in a room with three psychologists (something that sounds like a bad joke) but who also had incredible patience guiding me through the mathematical and computational modelling aspects of this project. These three luminaries delivered the

perfect measures of inspiration, challenge and good humour—I will truly miss our meetings together.

I would also like to take the opportunity to thank the University of Exeter Psychology department, and the SEORG research group, for providing me with a comfortable and stimulating working environment in which to conduct my research. I would like to extend a huge thank you to Karen Swanston for always being so friendly and helpful (and for providing me with the information no one else seems to know). I am also hugely appreciative to Rachel Arthur, for sharing her expertise in visual data analysis, and for being an interrater for this research project. I am grateful to Lancaster City Council and Barrow Borough Council for providing the CCTV footage that made so much of this empirical work possible. Thank you also to Lasse Liebst, Marie Rosenkrantz Lindegaard, Wim Bernasco, Don Weenink, Kim Møller, and all other members of the Interactionist Analyses of Violence Workgroup, for all of their valuable feedback on my research. I am grateful to the Engineering and Physical Sciences Research Council for financially supporting this research and its dissemination.

Throughout the PhD process, I have appreciated the camaraderie and friendship of the South Tower Crew. Therefore, I would like to extend my warmest thanks to Denise, Felicity, Matt, Melika, Merve, Modi, (lots of M's), Natasha and Tina. Tina Lina Agnes Frieda Keil (yes all just one person), you honestly need an entire page of dedication! Not only because your name is unusually long, but because you are an unusually wonderful person. You have been a true friend, my best friend, and you are a selfless angel—what would I do without you?... drink less coffee probably.

This thesis would not have been possible without all of the remarkable support, from all of these remarkable people, and for this, I will always be grateful.

Dedicated to my

little Kimpots

Abstract

Episodes of aggression and violence continue to beset our public spaces. This thesis explores how well we understand the transition to violence—and how aggression and violence in public spaces can be managed or controlled. We begin by arguing that established social psychological approaches to aggression and violence are inadequate for the task. Existing models explain violence through the failure of individuals to inhibit their own impulses or control their own emotions sufficiently. At best the models allow for the importance of dyadic interactions as individuals provoke each other as part of an escalation cycle. We argue that public space aggression and violence involves multiple parties and more complex sets of social dynamics. We suggest that, at the very least, the roles of third-parties and social categories need to be at the heart of theorising about violence in public spaces.

To support our arguments, we examined violence directly through detailed behavioural microanalyses of real-life aggressive incidents captured on CCTV footage. We also built agent-based models (ABM) to explore different theoretical approaches to the impact of groups and third-parties on aggression and violence.

The thesis contains seven studies. We begin with a CCTV behavioural microanalysis (Study 1) that showed collective group self-regulation of aggressive and violent behaviour in both within- and between-group conflicts. This study demonstrated an ‘intergroup hostility bias’, showing a greater likelihood of aggressive, escalatory acts towards outgroup members in intergroup conflicts than towards ingroup members in intragroup conflicts. Furthermore, this study demonstrated an ‘intragroup de-escalatory bias’, showing a greater likelihood of peace-making, de-escalatory behaviours towards ingroup members in intragroup conflicts than towards outgroup members in intergroup conflicts. Overall, we found that the majority of coded actions were acts of de-escalation performed by third-parties.

With evidence stressing the importance of social dynamics, we compared dyadic models of aggression against an alternative social model (which allowed normative influence

of others) in a dynamic agent-based modelling environment. We modelled the dynamics of metacontrast group formation (Studies 2 and 3), and found that group processes can produce both escalation of violence and inhibition of violence (Study 4). We found greater polarisation of violent positions in intergroup interactions than in intragroup interactions (Studies 5a and 5b). However, an emergent intergroup hostility bias did not emerge from this polarisation process. In Study 6, we re-examined the intergroup hostility bias present in our CCTV footage. We found an intergroup hostility bias for non-physical escalatory acts but not for physical escalatory acts. We examined the standardised number of actions contributed by third-parties and assessed the relationship between specific third-party conflict management strategies (policers and pacifiers) and conflict violence severity (Study 7).

Overall, our results showed that third-parties and groups are integral features of the dynamics of violence. Third-parties largely attempt to de-escalate conflict, and the conflict management strategy they employ has a direct relationship to the violent outcome. Groups have a tendency to de-escalate their own members, and self-policing and collective inhibition take place. These findings have importance for current models of aggression and violence and also for evidence-based violence reduction initiatives.

Table of Contents

Acknowledgements	ii
Abstract	v
List of Tables	xi
List of Figures	xii
Thesis Aims and Overview.....	14
1 Literature Review	19
1.1 Definitions of Aggression and Violence.....	20
1.1.1 Aggression as a Construct.....	20
1.1.2 Violence as a Physical Representation of Extreme Aggression.....	22
1.1.3 Social Psychological Explanations of Aggression and Violence—The Individual Balance Metaphor.....	23
1.2 A Critique of Established Models of Aggression	27
1.3 Microsociology, the Interaction Ritual and the Energy Metaphor	31
1.4 Violence in Primatology—Economics and the Value of Relationships	32
1.5 Beyond Inhibition—Bystanders, Guardians and Third-Party Policing.....	33
1.5.1 Psychology and Bystander Behaviour	34
1.5.2 Sociology and Guardianship	36
1.5.3 Third-parties in Non-Human Primate Communities.....	37
1.6 Social Group Categorisation and Its Effect on Aggressive Interactions	38
2 Analysing Violence in the NTE: A methodological review and systematic behavioural analysis of intergroup-intragroup violent interactions	46
2.1 Established and Emerging Methodological Approaches.....	47
2.1.1 Official Records	47
2.1.2 Interviews.....	48
2.1.3 Laboratory Settings	49
2.1.4 Naturalistic On-Site Observations	50
2.1.5 Video Analysis as an Observational Method	53
2.2 Group Behaviour in NTE Violence—New Insights and Avenues for Future Research Using CCTV	54
2.2.1 CCTV Limitations.....	58

2.3	Study 1: A CCTV Microanalysis of Within-Group and Between-Group Acts During Aggressive Encounters	61
2.3.1	Hypotheses	68
2.3.2	Method	69
2.3.3	Results.....	74
2.3.4	Discussion	80
3	Social Group Formation and Prototypical Positions in an Agent-Based Modelling Environment.....	89
3.1	Agent-Based Modelling	90
3.2	Agent-Based Modelling in Social Psychology.....	92
3.3	Study 2: Metacontrast and Social Group Formation in an Agent-Based Modelling Environment.....	95
3.3.1	The Model: Purpose, State Variables and Scales	95
3.3.2	Process Overview and Scheduling.....	98
3.3.3	Design Concepts	100
3.3.4	Initialisation, Inputs and Submodels.....	101
3.3.5	Simulation Experiments.....	101
3.3.6	Descriptions and Results of Scenarios	102
3.3.7	Discussion	106
3.4	Study 3: Self-Categorisation, Social Group Formation and Physical Flocking.....	108
3.4.1	Model: Purpose, State Variables and Scales	108
3.4.2	Process Overview and Scheduling.....	110
3.4.3	Design Concepts	113
3.4.4	Initialisation, Inputs and Submodels.....	114
3.4.5	Simulation Experiment	115
3.4.6	Results.....	115
3.4.7	Discussion	123
4	Dyadic and Social Agent-Based Models of Violence: The Roles of Dynamic Collective Inhibition and Polarisation in Group Violence	127
4.1	The Model: Purpose and Brief Description.....	128
4.1.1	Purpose.....	128
4.1.2	Brief Description.....	128

4.1.3	State Variables and Scales.....	129
4.1.4	Process Overview and Scheduling.....	132
4.1.5	Physical Space: Determining Social Groups and Movement	132
4.1.6	Social Space: Determining the Normative Context	133
4.1.7	Dyadic Space: Determining One’s Level of Aggression	135
4.1.8	Model Steps	137
4.1.9	Design Concepts	138
4.1.10	Initialisation, Inputs and Submodels.....	139
4.2	Study 4: A Comparison of Social and Dyadic Models of Violence	139
4.2.1	Hypotheses.....	141
4.2.2	Method	141
4.2.3	Results.....	142
4.2.4	Discussion.....	151
4.3	Study 5a: Investigation of Intergroup Hostility in Social Model of Violence.....	153
4.3.1	Hypotheses.....	154
4.3.2	Method	154
4.3.3	Results.....	154
4.3.4	Discussion.....	157
4.4	Study 5b: A Further Examination of Polarisation and the Effect on Violence.....	161
4.4.1	Method	161
4.4.2	Results.....	162
4.4.3	Discussion and ABM Evaluation	164
5	A Secondary Re-Examination of Intergroup Hostilities and the Role of Third-Party Involvement in NTE Violence	169
5.1	Study 6: An Extended CCTV Microanalysis of Within- and Between-Group Aggressive Acts and Level of Bystander Participation.....	170
5.1.1	Hypotheses.....	171
5.1.2	Methods.....	172
5.1.3	Results.....	172
5.1.4	Discussion.....	182
5.2	Study 7: Policers and Pacifiers—A New Typology of Third-party Intervention in NTE Conflicts	190

5.2.1 Hypotheses.....	195
5.2.2 Methods.....	196
5.2.3 Results.....	197
5.2.4 Discussion.....	201
6 General Discussion.....	206
6.1 Purpose of the Thesis	207
6.2 Brief Summary of Results	207
6.3 Theoretical and Practical Implications	210
6.3.1 Intragroup De-Escalation and the Collective Inhibition of Violence.....	210
6.3.2 Third-Party Bystander Conflict Management Strategies and Their Relationship to Violence Severity.....	214
6.3.3 Policy and Policing Implications	216
6.3.4 Final Evaluation of the Agent-Based Model.....	217
6.4 Violence as Choreography: Future Trajectories	223
6.5 Concluding Summary.....	227
References.....	229
Appendix	263
Video Coding Book	263

List of Tables

Table 2.1	The likelihood of the next sequential action succeeding an act of intragroup escalation, intragroup de-escalation, intergroup escalation and intergroup de-escalation, ($N = 3512$)	80
Table 3.1	Results from a binomial logistic regression predicting migration from a social group.....	123
Table 4.1	Moran's I auto-spatial correlation for X-Y coordinates and PoV for two models.....	146
Table 4.2	Comparison of ICC(1)s depicting homogeneity within groups on PoV for the dyadic and social models	148
Table 4.3	Comparison of ICC(2)s depicting heterogeneity between groups on PoV for the dyadic and social models	149
Table 4.4	Comparison of ICC(1)s depicting homogeneity within groups for PoV and prototypical	150
Table 5.1	The likelihood of the next sequential action succeeding an act of intragroup/intergroup physical escalation, intragroup / intergroup non-physical escalation and intragroup/intergroup de-escalation, ($N = 3511$)	180
Table 5.2	Bivariate correlations of bystander role measures and severity of violence.....	197
Table 5.3	Hierarchical multiple regression analysis of the number of bystander, pacifiers and policers in predicting the severity of violence	199

List of Figures

Figure 1.1	General Aggression Model	25
Figure 1.2	Representation of the I ³ theory	26
Figure 1.3	Violence escalation cycle.....	29
Figure 2.1	Examples of an escalatory behaviour in the form of a punch (left) and a de-escalatory behaviour in the form of holding people apart (right).....	57
Figure 2.2	Total number of de-escalatory and escalatory acts as a percentage of the sum of all behaviours ($N = 3555$)	74
Figure 2.3	Total number of intragroup and intergroup acts as a percentage of the sum of all behaviours ($N = 3555$).....	75
Figure 2.4	De-escalation as a proportion of intergroup total acts and intragroup total acts for each video ($N = 74$).....	77
Figure 2.5	Escalation as a proportion of intergroup total acts and intragroup total acts for each video ($N = 74$).....	78
Figure 3.1	Model of Segregation	91
Figure 3.2	Example of NetLogo in-radius neighbourhood	98
Figure 3.3	Prototypicality curves for physicists and biologists	103
Figure 3.4	Investigators' prototypicality curves for physicists and biologists.....	104
Figure 3.5	Prototypicality curves for physicists, biologists and social scientists	105
Figure 3.6	Investigators' prototypicality curves for physicists and biologists, and social scientists	106
Figure 3.7	An example of connected flocking groups	112
Figure 3.8	Simulation at iteration zero.....	116
Figure 3.9	Simulation at iteration 3588.....	116
Figure 3.10	Ripley's K at the start of the simulation	117
Figure 3.11	Ripley's K test after 3588 iterations	118
Figure 3.12	Modelling example of group merging and separation	121
Figure 4.1	Social model of violence	129
Figure 4.2	Linear function mapping agent aggression to PoV.....	137
Figure 4.3	Dyadic model of violence	141
Figure 4.4.	Tick zero—Norms constant	143

Figure 4.5	Tick zero—Norms update social model. Shaded by heading	143
Figure 4.6	Iteration 4267—Norms constant dyadic model	145
Figure 4.7	Iteration 4267—Norms updated social model.	145
Figure 4.8	Polarisation of positions on violence in intergroup and intragroup interactions	156
Figure 4.9	Colour social group of five agents with similar positions on violence.....	158
Figure 4.10	Colour social group of five agents with divergent positions on violence.....	159
Figure 4.11	Polarisation of positions on violence in intergroup and intragroup interactions.....	162
Figure 4.12	Bidirectional polarisation of positions on violence in intergroup and intragroup interactions	164
Figure 5.1	Proportions of de-escalatory, physical escalatory and non-physical escalatory acts as a percentage of the sum of all behaviours, ($N = 3555$)	173
Figure 5.2	A barycentric coordinate systems plot representing the proportion ratio of physical escalation, non-physical escalation and de-escalation for each video, ($N = 74$).....	174
Figure 5.3	The average proportion of intragroup de-escalation, physical escalation and non-physical escalation across the 43 cases.....	176
Figure 5.4	The average proportion of intergroup de-escalation, physical escalation and non-physical escalation across the 31 cases.....	176
Figure 5.5	The standardised number of actions by role, N protagonists = 43, N targets = 43, N active bystanders = 242	182

Thesis Aims and Overview

Public space violence can result in physical, psychological and societal harm—all of which burden public services and concern communities. The rise of the night-time economy (NTE)—characterised by leisure zones, cheap alcohol deals and extended drinking hours—is associated with increases in reported violence and antisocial behaviour (Hadfield, Lister, & Traynor, 2009; Hobbs, Lister, Hadfield, Winlow, & Hall, 2000). The latest UK Security Industry Authority statistics reflect this rise and show that up to 50 % of all violent incidents take place within spatial proximity of NTE zones, with more conservative estimates suggesting that one in five violent incidents occur in and around drinking establishments (Porter, 2015). Unsurprisingly, as many as 75 % of NTE patrons have witnessed violent encounters, with 10 % – 15 % having a direct involvement in a conflict (Porter, 2015). This increase in public space violence, which is depicted as a crisis for state policing, has led to greater public anxiety and to new political reforms (Bromley, Thomas, & Millie, 2000; Hobbs, Hadfield, Lister, & Winlow, 2005; Lister, 2009). The broad aim of this thesis is to combine social-psychological theory with cutting-edge technological techniques—such as CCTV behavioural analysis of real-life violence in public spaces and computational modelling of theoretical models of aggression—to better understand, predict and prevent violence in public spaces.

Established models of aggression and violence in social psychology—such as the general aggression model (GAM) and the I³ theory (Anderson, 1997; DeWall, Anderson, & Bushman, 2011; Finkel, 2007, 2014)—have identified an impressive list of personal and situational risk factors that are relevant in determining the likelihood and strength of aggressive impulses. However, the dynamics of violence escalation in these individualist accounts are restricted to the provocations received by another and the individual capacity to incite or inhibit one's own aggressive response. Therefore, at best, these traditional approaches explore violence as a dyadic relationship, and exclude the wider social context

and more complex social dynamics between multiple actors. This emphasis on the aggressive dyad is especially problematic in the domain of public space violence that most frequently occurs in the presence of an audience and is often group-based (Levine, Taylor, & Best, 2011; Liebst, Heinskou, & Ejbye-Ernst, 2017; Wells & Graham, 1999).

This thesis is an attempt to address the relative neglect of social groups and social contexts as important explanatory components in social-psychological models of aggression and violence. Specifically, it explores how groups and third-parties may shape the trajectory of violence beyond the aggressive dyad. Furthermore, it addresses the effect of social group membership on rates of de-escalatory and antagonistic behaviours in public space violence.

Chapter 1 provides a theoretical justification for the project and examines the available definitions of aggression and violence. We review the psychology, sociology and the primatology violence literatures, and scrutinise how well current theory understands the emergence of violence. We summarise the dominant models of aggression in social psychology (such as the GAM and I³ theory) that tend to focus on the capacity of the individual to inhibit their own propensity for violence. We then consider alternative strands of literature that emphasise the importance of third-parties and social group categorisations as explanatory components of the emergence and regulation of violence. Building upon these literatures, we argue that it is necessary to go beyond dyadic interactions to fully understand the occurrence of violence.

Chapter 2 is a methodological review of the empirical methods chiefly utilised by those studying public space violence, with particular emphasis on recent theoretical contributions from CCTV analysis studies. We outline how extensive CCTV coverage of public spaces can be an exceptional resource for detailed behavioural analyses of real violent interactions. We further describe how this emerging method has provided new insights into the nature of violence and allowed researchers to challenge the orthodoxy in the psychological literature, in which group processes are seen to create the conditions for emerging violence.

Simultaneously, this method has allowed exploration of the patterns of social regulation of violence predicted by sociological theory, but which have hardly been studied directly by sociologists. Following this methodological review, we present our first empirical study that extends previous behavioural CCTV analyses by examining the role of social group membership on rates of escalation and de-escalation in intragroup and intergroup public space conflicts. We find that the majority of all actions performed are by bystanders and that these acts tend to be de-escalatory. We record an intragroup de-escalatory bias, whereby there is a greater likelihood of de-escalatory peace-making behaviours towards ingroup members in intragroup conflicts than towards outgroup members in intergroup conflicts. Furthermore, we find an intergroup hostility bias, whereby there is a greater likelihood of intergroup-directed aggressive escalatory actions during intergroup conflicts than intragroup-directed escalatory actions during intragroup conflicts. In light of these group-based differences, we revise established models of aggression to take into account social groups and local social context. We employ agent-based modelling to dynamically and ethically compare established dyadic accounts of aggression and violence with an alternative model that encompasses social considerations.

The main aim of Chapter 3 is to computationally model social group formation and social context in a manner that can then be incorporated into a broader agent-based model (ABM) of violence. The chapter begins with a brief description of agent-based modelling and how it has been used in social psychology. We then develop our ABM of violence by modelling the dynamics of metacontrast group formation as formulated by Salzarulo (2004, 2006). In Study 2, we conduct replica experiments of Salzarulo's (2004) original work to demonstrate the robustness and validity of our adapted prototypicality function. In a novel experiment (Study 3), we employ the principle of metacontrast to form context-dependent social groups which move together in a physical space. Analysis of the make-up of these social groups shows that agents are not spatially clustered by chance; rather, those more

similar on opinion dimensions are more likely to share social groups and be closer together in physical space. Being psychologically distant from a social group prototype, rather than physically distant from the centre of the social group, predicts agent migration.

Chapter 4 tests the principles of traditional psychological approaches to violence—represented by the principles of I³ theory—against an alternative approach that builds on the social group modelling developments from Chapter 3. We computationally formalise I³ theory into an agent-based modelling environment and compare this model with our social identity-based model in which individuals derive prototypical positions on violence from the dynamic local social context (Studies 4, 5a and 5b). We find greater spatial clustering of violence and nonviolence in the social model than in the established model. We further record stronger intra-group consistencies of violence in the social model than in the established model. Finally, we find that social groups significantly differ more in their levels of violence in the social model than in the dyadic model. As such, individuals in the social model can be inhibited or uninhibited by the local social context. In Studies 5a and 5b, we investigate whether a self-categorisation-informed model of violence can produce the intergroup hostility biases predicted by self-categorisation theory (e.g., Otten, 2009; Viki & Abrams, 2013) and demonstrated in Study 1. We find evidence of greater polarisation in intergroup interactions as opposed to intragroup interactions. However, we find no evidence of an emergent intergroup hostility bias through this polarisation mechanism.

In Chapter 5, we conduct a secondary data analysis of the CCTV footage coded in Study 1. The aim of this extended work is to provide more refined information extracted from the footage that can then be used for future ABM developments. A further objective is to refine our understanding of intragroup and intergroup behavioural distinctions as observed in our CCTV analyses. Therefore, in Study 6, we examine intragroup and intergroup behavioural differences when unpacking escalation into physical and non-physical behaviours. We find evidence of an intergroup hostility bias for non-physical escalation (in line with CCTV

Study 1) and find no evidence of differences in levels of physical escalation between intragroup and intergroup interactions (in line with ABM Studies 5a and 5b). The aim of our final study (Study 7) is to dissect the different bystander roles observed in our CCTV footage and examine whether third-party intervention is a division of labour with degrees of specialisation that can impact on violence severity. We find that the number of active third-party bystanders is negatively associated with violence severity. However, this relationship is negated once accounting for those bystanders who de-escalate both antagonists—referred to as policers. On the other hand, the presence of bystanders who de-escalate only one of the antagonists—pacifiers—is not predictive of violence severity.

Overall, the results reported in this thesis suggest that third-parties and social group categorisations are integral features of the dynamics of violence in NTE conflicts. Third-party bystanders largely try to de-escalate conflict and the conflict management strategy they employ has a direct relationship with the violent outcome. Third-parties are important for understanding conflict escalation and de-escalation; as too are the shared and distinct social categories to which individuals belong. Specifically, escalatory and de-escalatory behaviours are not randomly directed, but are highly socially bound. What is most evident is that groups have a tendency to de-escalate their own members, and that self-policing and collective inhibition is taking place.

In our final chapter (Chapter 6), the implications of these findings, the avenues for future research and an assessment of our ABM of violence are discussed.

1 Literature Review

Violence is a persistent societal concern that threatens communal living and individual quality of life. This literature review examines how well current theory understands the emergence of violence in human societies. More specifically, it compares the explanations for the emergence of violence in the human and non-human primate literatures. We argue that each literature orientates to the problem through a different organising metaphor, with a different emphasis placed on the importance of third-parties.

We show that researchers interested in the psychology of human violence have tended to focus on the capacity of the individual to inhibit their own propensity for violence (Anderson, 1997; DeWall et al., 2011; Finkel, 2007, 2014). This may be a failure of emotion regulation or cognitive control, or the culmination of an antagonistic spiral in which one partner in a dyad fails to inhibit an unnecessarily violent response (Anderson & Carnagey, 2004; Slotter & Finkel, 2011). As such, psychologists see violence through the lens of a balance metaphor, where loss of control is the result of a loss of balance between resources that support control and those that undermine it. Conversely, sociologists of violence are less interested in the characteristics of the individual protagonists and focus instead on the aggressive interaction (Collins, 2008, 2009). Here, confrontational tension and fear provide natural barriers that inhibit a violent response. Violence emerges when the emotional energy associated with the conflict finds a way around the natural barriers to the expression of violence. As such, the transition into violence is seen through the lens of an energy metaphor. Finally, we demonstrate how researchers of non-human primate violence tend to focus on the quality of the relationship between antagonists and the potential cost for the individual in engaging in violent acts (de Waal, 2000; Georgiev, Klimczuk, Traficonte, & Maestripiერი, 2013). Here, violence emerges from the failure of ritual displays and a calculation that the rewards of violence exceed the risk of engagement. As such, the transition to violence is seen through the lens of an economic metaphor.

Building upon these literatures, we argue that it is necessary to go beyond dyadic interactions, and beyond valued relationships, to fully understand the occurrence of violence. We make this claim because most violence occurs in the presence of third-parties, and the importance of third-parties to violence is not adequately recognised in any of the main theoretical approaches. We propose that third-parties are more than a generic feature of the ambient context in which violence emerges from an aggressive dyadic interaction. Third-parties are actual (or potential) participants in the stream of aggressive interactions that precede and succeed a violent outbreak. Woven within and between all relevant actors are relevant social categories that provide expectations and give meaning to uncertain events.

Before we explore the role of third-parties in the transition of violence, it is necessary to review the definitions of ‘aggression’ and ‘violence’.

1.1 Definitions of Aggression and Violence

1.1.1 Aggression as a Construct

Aggression is something that is easy to recognise and yet conceptually difficult to encapsulate. Superficially, one may perceive aggression as a manifestation of spilled emotion, behaviourally expressed in destructive actions intended to harm. Yet, with further consideration, the word ‘aggression’ is pragmatically applied to a vast range of behaviours that, although sharing similar properties, remain distinct. Early conceptualisations of aggression focused on physical interactions. Ethologists Scott (1958) and Lorenz (1966) were parsimonious in their definition of aggression as an internal fighting instinct expressed through physical attacks on other members of the same species. This conceptualisation was criticised for the emphasis that was placed on physical assault—a case put forward by Rummel (1979), who identified that aggression could be many things: “we can cast an aggressive eye at a party, invade a person’s quiet, or attack another verbally without threatening or inflicting physical harm or injury” (p. 202). While ethologists arrived at a

rather narrow definition of aggression, psychologists of the 1960s defined aggression more broadly as any “response that delivers a noxious stimuli to another organism” (Buss, 1961, p. 3). While this definition subsumes the many incarnations of aggression, both physical and psychological, it is still a definition that fails to look beyond overt behaviour. For Kaufmann (1965) and others (Berkowitz, 1969; Goldstein, 1975), aggression also had to “have a subjective probability > 0 —on the part of the attacker—of reaching that object, and either removing it from the attacker’s goal path, or imparting a noxious stimulus to it, or both” (Kaufmann, 1965, p. 353). To put it plainly, there must be intentionality to harm (anticipation) regardless of whether or not the noxious stimuli is delivered. By incorporating intentionality into the definition of aggression, acts that unintentionally cause damage (such as accidentally hurting another person) are no longer classified as aggressive. Conversely, acts that intend to harm but that do not reach the target (e.g., a punch thrown that misses the intended target or a malicious rumour spread about an individual who is unaware of the stories being spread) are classified as aggressive. Therefore, aggression can be understood in terms of an underlying motivation to harm another being, as opposed to a specific action that successfully harms an intended target.

Although the inclusion of intentionality is sensible, this creates certain complexities that make classifying acts of aggression more ambiguous. Essentially, while the intention to harm another may be observable, motivations may need to be inferred. However, assessing one’s intentions is largely subjective and, thus, outside the domains of rigorous analysis (Geen, 1990). Despite these added complexities, Geen (1990) proposed that inferences regarding the intent to harm were inescapable, and he included them in his definition of aggression that remained dominant throughout the 1990s. For Geen (1990), aggression was defined as “the delivery of a noxious stimulus to another person with the intent of harming that person, and in the expectation that the aversive stimulus will reach its destination” (p. 28). While this definition was largely accepted as sufficient, others argued that such a

conceptualisation ignored acts of harm that were inflicted with the target's consent (e.g., acts of sadomasochism and medical procedures). Baron and Richardson (1994) attempted to distinguish these consented acts from aggression by defining aggression as "any form of behaviour directed toward the goal of harming or injuring another living being who is motivated to avoid such treatment" (p. 7). This definition, which includes underlying motivations to harm (intention) and excludes consented acts of harm, remains the most predominant conceptualization of aggression in current literature, and has been widely applied to a vast array of research domains. However, while this definition is attentive, it remains dependent on the idea that aggression is always something destructive. For other researchers, aggression in certain settings (such as in business environments and in the school playground, in self-defence and through collective action, during competitive events or in military contexts) can be desirable, socially acceptable and even adaptive (see Little, Rodkin, & Hawley, 2013). Finally, while an aggressive act itself may be relatively easy to recognise, the motivations to instigate such behaviour may vary significantly (see Geen, 1990, for an overview of affective/reactive aggression and instrumental/proactive aggression).

1.1.2 Violence as a Physical Representation of Extreme Aggression

To this point, it is clear that aggression is a multifaceted construct encompassing a plethora of behaviours. Aggression can be defined narrowly and, as such, may exclude referents relevant to the concept. Alternatively, aggression can be broadly defined to include underlying motivations, which introduces ambiguity. Such semantic disputes remain when considering the term 'violence'. Definitions of violence, similar to those of aggression, are far from standardised and vary as a result of their political, social and psychological applications (Tolan, 2007). Much the same as aggression, violence is broadly defined to include both intentionality to harm and non-physical acts, such as coercion, intimidation and threat (Tolan, Gorman-Smith, & Henry, 2006). However, the distinction between the terms 'aggression' and 'violence', for most researchers, lies in the extremity of the behaviour (Anderson & Bushman,

2002; APA, 2015; DeWall et al., 2011; Felson, 2009). Aggression harbours a spectrum of behaviours, and violence represents a subset of behaviours that are most severe on the aggression continuum. As such, all violent acts are aggressive, but not all aggressive acts are violent (Bushman & Huesmann, 2010). Returning to semantics—although violence may indeed be a representation of extreme aggression, it is not to say that violence is a smooth transition from extreme aggression. For the majority of people, the prospect of violence is met with a barrier of confrontational tension and fear making it difficult (Collins, 2008).

Given these considerations, the field remains at an impasse. Therefore, researchers have directed much attention into understanding the causes of aggression and the escalation into violence. For the purpose of this thesis, we restrict our working definition of violence to: any intentional physical act perpetrated by an individual towards another that inflicts physical harm on the intended target. The work of this thesis largely focuses on violence in NTE settings committed between individuals sharing this public space. These violent episodes tend to involve individual skirmishes, which may be informed by social groupings, but which are not ordered by a higher authority. Therefore, our current interests exclude forms of violence between larger, higher-ordered entities; including violence by terrorist groups, state-sanctioned inflicted violence, wars between countries and political violence, to name a few.

1.1.3 Social Psychological Explanations of Aggression and Violence—

The Individual Balance Metaphor

There is no single cause of aggression. Previous social psychological research has supposed that aggression may be a reaction to negative affect generated through an aversive event, such as frustration, physical pain, unpleasant noises or heat (neoassociation theory; Berkowitz, 1989, 1993; Dollard, Doob, Miller, Mowrer, & Sears, 1939), or a reaction to misattributed lingering physiological arousal (excitation transfer theory; Zillmann, 1979, 1988). Learning theories suggest that aggressive responses may be developed through the

perception of oneself or another attaining certain goals or rewards for behaving aggressively (social learning theory; Bandura, 1973, 1977) or the enactment of a well-learned aggressive mental representation (script) retrieved from semantic memory (script theory; Huesmann, 1988, 1998; Huesmann & Taylor, 2006). Alternatively, and less dependent on automaticity, individuals may be decision-makers who weigh up the outcomes of using aggressive behaviour as a means of fulfilling an instrumental goal—for example, obtaining money or information, righting a perceived wrong, and tailoring or maintaining an identity such as a tough reputation (social interaction theory; Tedeschi & Felson, 1994).

Incorporating these theories of aggression into a broader model, Anderson and colleagues (Anderson, 1997; Anderson & Bushman, 2002; Anderson & Carnagey, 2004; Anderson, Deuser, & DeNeve, 1995) proposed the overarching general aggression model (GAM) that considers both *personal* (e.g., gender, individual beliefs, attitudes, genetic predispositions, goals and personality traits) and *situational* (e.g., pain and frustrations, aggressive cues and provocations) input variables as risk factors of aggression. The model posits a three-stage process (see Figure 1.1) in which these personal and situational inputs interact to alter one's internal states, subsequently influencing *cognitions* (aggression-related thoughts, schemas and scripts), *affects* (moods and emotions) and *arousals* (physiological changes strengthening action tendencies). These three internal states then influence the perceiver's immediate appraisal of the situation that feeds into the behavioural choice the individual makes. Individuals who are short of sufficient motivation or resources to reappraise the incident are likely to behave aggressively, while those with greater motivation and resources are more prone to acting thoughtfully and in a reassessing manner (self-control was later added to the GAM by DeWall et al., 2011). The action taken then feeds back into the personal and situational inputs, thus restarting the cycle.

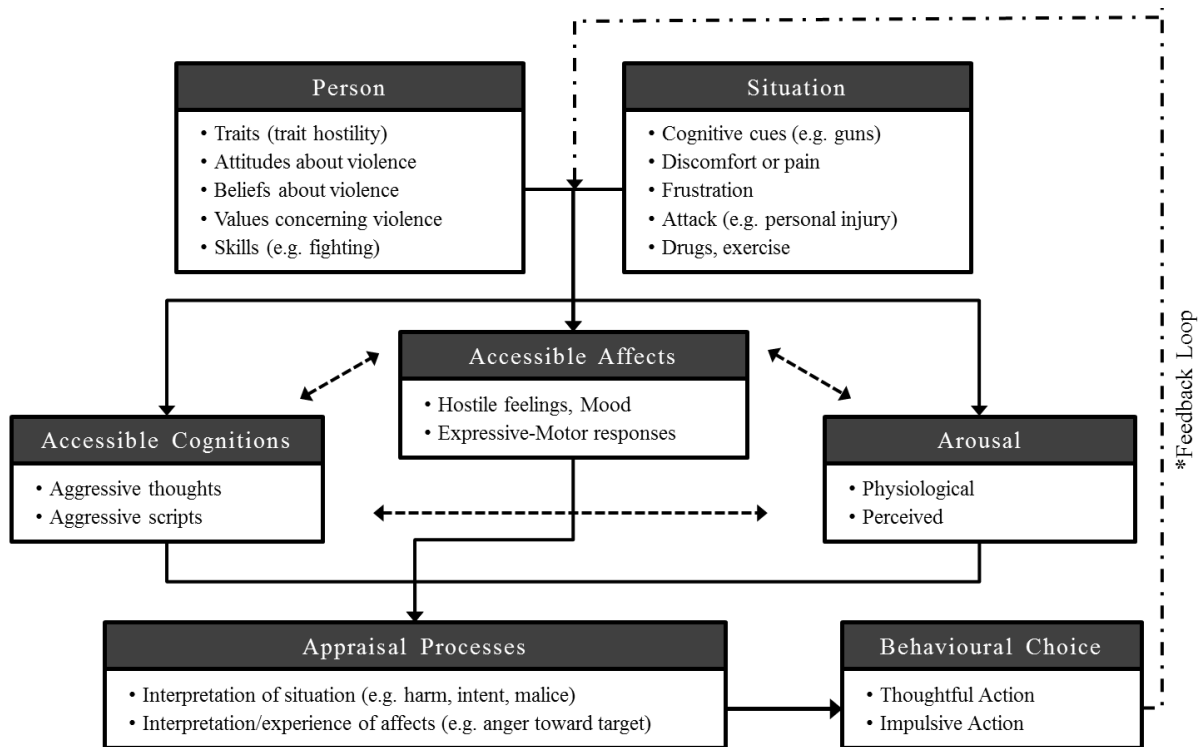


Figure 1.1. General Aggression Model (adapted from Anderson, 1997; feedback loop incorporated by Anderson & Bushman, 2002)

A more recent meta-theory of aggression, heavily influenced by the GAM, revealed how the concept of a balance metaphor is used to account for an individual loss of control. The I³ theory (pronounced ‘I-cubed’ theory) (Finkel, 2007, 2014; Finkel, Bodenhausen, & Bushman, 2008; Finkel & Eckhardt, 2013; Slotter & Finkel, 2011) reconceptualises key aspects of the GAM and represents the model as a balance scale (see Figure 1.2). The I³ theory introduces three processes—instigation, impellance and inhibition—that interact to determine the likelihood and strength of an aggressive response in each situation (see Figure 1.2). *Instigation* is the trigger of an aggression episode and refers to the perceiver’s experience of a direct or indirect provocation, insult, goal obstruction or social rejection. This provocation alters the perceiver’s arousal, cognitions and affect and combines with impellance to determine the ‘urge readiness’ to behave aggressively. *Impellance* is the ‘psychological preparedness’ to act aggressively in response to the instigation and consists of two subcategories: *dispositional* and *situational* impellance. These two categories map onto

the *personal* and *situational* input variables of the GAM and include dispositional attributes deriving from the individual (such as one's own trait aggressiveness, hostility, rumination and narcissism) and situational factors existing from the environment (such as noise, crowding, temperature and pain). The sum of the instigation and impellance is finally balanced against inhibition, which is the ability to inhibit one's aggressive urge. Inhibition derives from dispositional (e.g., self-control ability, executive functioning and traits) and situational (e.g., self-regulatory resources, low levels of alcohol or drug use) factors. When the aggressive urge is stronger than one's inhibition, the person reacts to the provocation with an aggressive response—the severity of which increases as instigation and impellance is high and inhibition is low ('perfect storm' phenomenon; see Finkel, 2014, for experimental review).

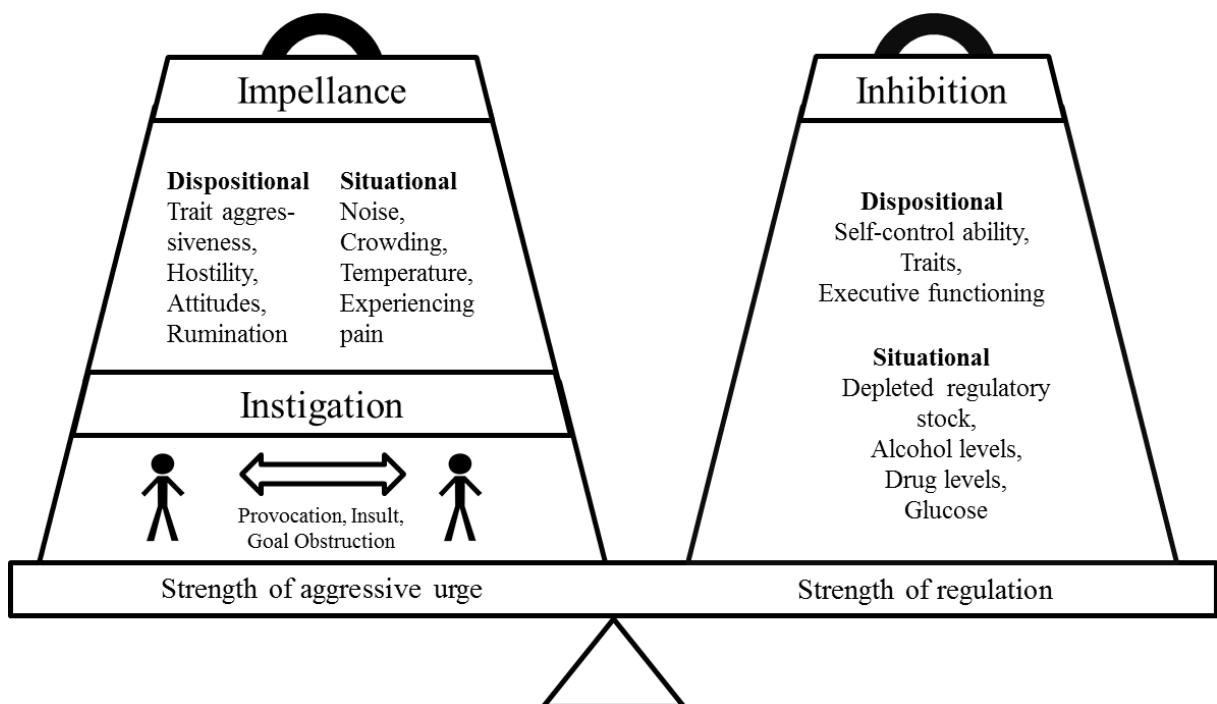


Figure 1.2. Representation of the I³ theory

1.2 A Critique of Established Models of Aggression

These established models of aggression have identified an impressive list of personal and situational risk factors that are relevant in determining the likelihood and strength of aggressive impulses. However, these paradigms fundamentally model mild laboratory aggression and are less effective in describing or predicting real-world aggression (see Ferguson & Dyck, 2012).

One explanation for this failure originates from the experimental operationalisations of aggression used to support further violence theory. These paradigms typically measure aggression using dependent variables such as hot sauce doses, mild shock or noise blast administration, written retaliatory insults, negative feedback scores and even virtual pins in virtual voodoo dolls (for examples, see Finkel, 2014). Although novel approaches are required owing to ethical restrictions, it remains debatable whether these behavioural responses (which may be inadvertently cued by the experimenter; Ritter and Eslea 2005) are in fact aggressive, violent or perhaps even amusing. Results from these studies have been applied to shape our understanding of violence across multiple domains, from intimate partner violence to racial genocides and suicide to torture (Anderson & Carnagey, 2004; Benjamin, 2008; DeWall et al., 2011; Finkel & Eckhardt, 2013). This provides researchers with a ‘one-size-fits-all’ model of violence that ultimately ignores “the complexity, diversity, and contextual aspects of real-world aggression” (Graham et al., 2006).

Another explanation for the failure to predict real-world violence is the interchangeable way in which authors drift between the terminology of ‘aggression’ and ‘violence’. For example, when defining their own I³ theory, Finkel et al. (2012) stated: “when the strength of inhibition exceeds the strength of the urge to *aggress*, people behave *non-violently*; when the reverse is true, they behave *violently*” (emphasis added) (p. 534). For the authors, their model of aggression is interchangeably used to explain violence without taking any additional considerations into account. Additionally, in this example, violence is not even

extreme aggression, but any response beyond the inhibition threshold. Are we to determine from this that after exceeding inhibition, an aggressive word spoken to another is considered violence?

Even when violence is facilely defined as extreme aggression, attempts to extend the GAM to explain violence (e.g., the review of DeWall et al., 2011) have been theoretical, scant in terms of empirical evidence, and largely insufficient (for a critique, see Ferguson & Dyck, 2012). One such attempt is the GAM–influenced Violence Escalation Cycle (VEC) (Anderson & Carnagey, 2004), which assumes that violence is the result of escalating retaliation between two parties (e.g., two people, two groups or two nations). At the offset, the cycle is triggered by a mild to major event (see Figure 1.3). Person A reacts to this event by harming person B in a way they feel is appropriate and justified. Person B may consider this action unintentional, justified or relatively mild and not react, thus ending the cycle. Alternatively, person B may consider the action intentional, unjustified and relatively harmful and react with heightened aggression. Person A then evaluates the response as either appropriate retaliation and ends the cycle, or inappropriate over-retaliation that warrants a heightened aggressive response. This cycle continues and escalates into violence.

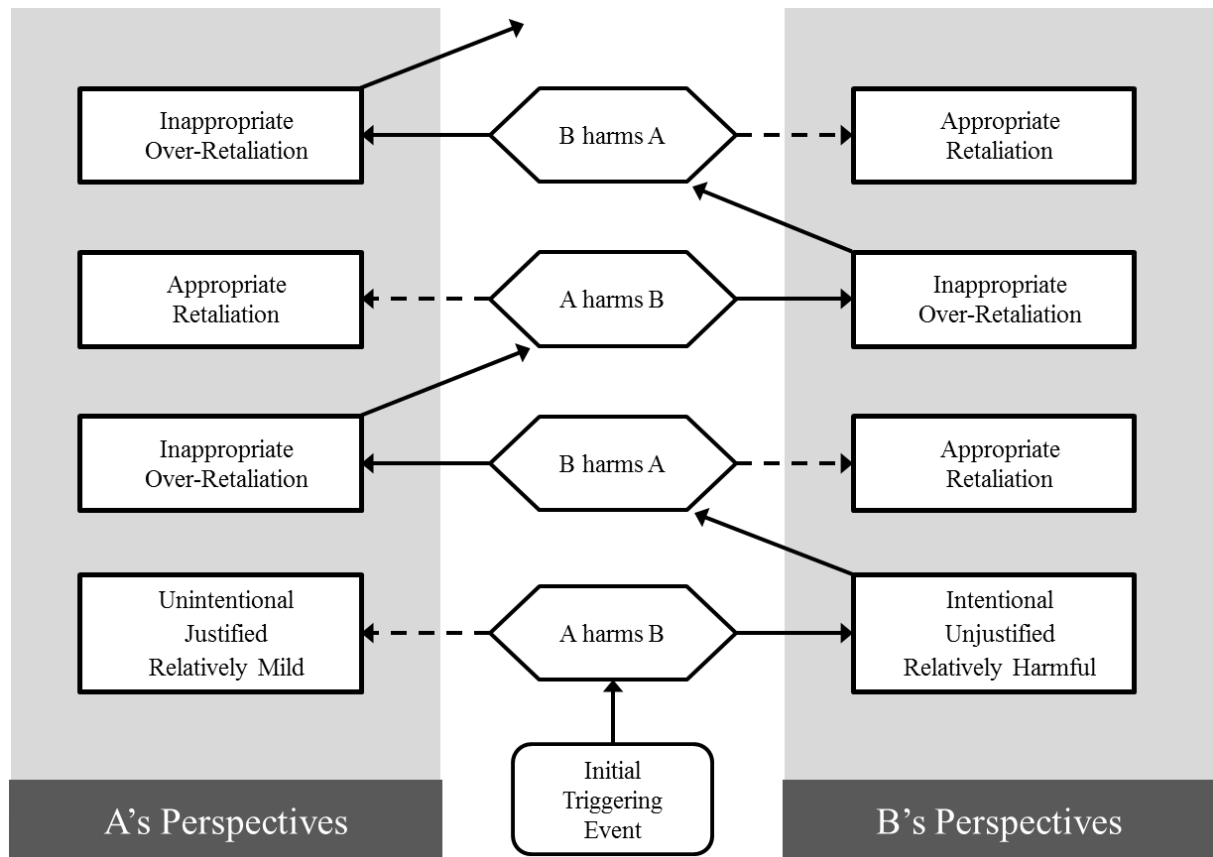


Figure 1.3. Violence escalation cycle (adapted from Anderson & Carnagey, 2004)

While violence is often referred to as an escalatory process (a brief Web of Science search of ‘violence escalation’ yields 497 results, search date 7 December 2016), this cycle seems oversimplified and largely ignores the personal and situational risk factors of aggression offered by the GAM framework (see Figure 1.1).

It is apparent from the most recent GAM, I³ theory, and VEC that violence is understood as being the result of a failure of inhibitory balance. This balance is located at the intrapsychic level and, as such, loss of personal control may tip the perpetrator into violence. However, the models have very little to say about the role of the relationship between antagonists (other than that they are the source for potential provocation) or the role of third-parties (other than that they may be part of the situational Impellence in an unspecified way). Therefore, a final, and chief, criticism of these models is their overemphasis on the capacity of the individual to inhibit their own aggressive urges at the expense of the dynamic wider

social context. Specifically, although violence may be an interaction between two people in isolation, it more often occurs in the presence of an audience (Felson, 1982; Wells & Graham, 1999), who have the capacity to shape and alter both the immediate situation and the potential violent outcomes (Levine, Lowe, Best, & Heim, 2012; Levine et al., 2011).

In support of this point, figures collected by the United States (US) Department of Justice demonstrated that the presence of third-parties was indeed characteristic of human violence. A US Bureau of Justice Statistics (BJS) report concluded that third-parties were present during two-thirds of all violent victimisations in the US between 1993 and 1999 (Planty, 2002). Third-parties were present at 70 % of assaults, 52 % of robberies and 29 % of rapes or sexual assaults. Even domestic violence, often imagined as a form of violence hidden from the eyes of others, had a significant chance of being conducted in the presence of third-parties. The BJS figures revealed that around one-third of all intimate partner violence occurred in the presence of a third-party compared with two-thirds of all violent acts that occurred between strangers or other acquaintances. This figure is believed to be even higher in the context of public space violence, where third-party presence is ubiquitous (Levine et al., 2011; Liebst et al., 2017; Wells & Graham, 1999).

We will shortly put forward the case that third-parties are more than a generic feature of the ambient context in which violence emerges from an aggressive dyadic interaction. Third-parties can intervene actively or observe passively (recognising that inaction is often understood as a form of action by antagonists). They can intervene on behalf of one party or another. They can interact with one party or both. They can contribute behaviours that seek to escalate or de-escalate the action. They can interact with other third-parties (rather than the antagonists) in ways that seek to escalate or de-escalate. Put simply, third-parties are an integral feature of the dynamics of violence and are equally as important as the capacity of individuals to inhibit their own violent impulses.

1.3 Microsociology, the Interaction Ritual and the Energy Metaphor

Alternative strands of research stress the importance of this wider social context in understanding and better predicting violence. For example, microsociologists have postulated that violence cannot be fully understood by assessing only the background factors of a perpetrator. Rather, violence is a social interaction that emerges from the immediate situation comprised of a series of micro-interactions between all relevant actors (Collins, 2008). Collins (2008) states:

Most existing explanations fall into the category of background explanations: factors outside the situation that lead up to and cause the observed violence. Some background conditions may be necessary or at least strongly predisposing, but they are certainly not sufficient; situational conditions are always necessary, and sometimes they are sufficient, giving violence a much more emergent quality, than any other kind of human behaviour. (p. 20)

The central idea of this proposition is that demographic variables such as gender, ethnicity, social class and individual dispositional traits fail to explain why those classified as 'high risk' remain passive during altercations, while those deemed less 'at risk' act violently. This failure results because violence is an *interaction* that *emerges* from the immediate situation that comprises a series of micro-interactions between all relevant actors. By default, humans—even trained soldiers—find violence difficult (Collins, 2009). This is explained by a neurologically hardwired propensity for interactional entrainment and solidarity with others (Collins, 2009). This propensity for entrainment and solidarity means that any violent urge is met with a barrier of confrontational tension and fear that, for the most part, holds violence at bay. However, certain pathways can convert confrontational tension and fear into an emotional energy that can circumvent inhibitory barriers and drive violence. One pathway is a sudden perception of weakness in the victim, which terminates any joined entrainment. A second pathway is the presence of a supportive 'stage' audience who can encourage a contest

by keeping activities within ‘socially enforced limits’ (Collins, 2008). In this account, the propensity for violence is not bound solely within the individual; rather, to attack another requires “reorganizing the emotions as an interactional process involving everyone present: the antagonists, audience, and even ostensibly disengaged bystanders” (Collins, 2008, p. 8). Placing aside the potential for victim blaming, and the rather undynamic conceptualisation of the role of the audience in violent encounters, the key claim is the importance of a natural inhibition that is written into human interactions. The path to violence is therefore understood as a failure of inhibition at the level of the collective interaction.

1.4 Violence in Primatology—Economics and the Value of Relationships

In the animal behaviour literature, violence occurs when the mechanisms that have evolved to reduce the risk of death and injury associated with conflict (e.g., ritualised fighting, dominance displays and signal exchanges) somehow fail. Therefore, when violence does emerge, it is usually seen as the result of an individual cost–benefit calculation that concludes that the benefits of engagement in violence exceed the potential costs entailed (Georgiev et al., 2013). In the non-human primate literature, this cost–benefit calculation has been augmented to include a social relationship component. Specifically, the vast majority of aggression in social animals involves familiar individuals who share a past and are expected to share a future (de Waal, 2000). In this context, violence can damage longer-term relationships that are themselves commodities with an attached value. De Waal (2000) points out that “in many social animals, however, both parties stand to lose if escalated fighting damages relationships” (p. 589). Survival depends on reciprocated assistance, and so aggression needs to be constrained to protect mutually beneficial relationships: “in other words, social relationships are commodities the deterioration of which needs to be prevented” (de Waal, 2000, p. 588). In this example, we can see both the economic metaphor in play (‘relationships are commodities’) and the importance of valued relationships as key impediments to the expression of violence. In the non-human primate literature, valued

relationships inhibit violence in the first instance, as well as facilitate post-conflict reconciliation in which former opponents console one another for relationship repair (Aureli, Cords, & van Schaik, 2002; Aureli & de Waal, 2000; de Waal, 2000). Therefore, the overarching theme of the primate literature on the transition to violence is the importance of valued relationships in individual cost–benefit calculations.

1.5 Beyond Inhibition—Bystanders, Guardians and Third-Party Policing

A common thread to be pulled from these three literatures is the general agreement that an inhibition mechanism operates to prevent violence. However, there is disagreement about where the inhibition is located. In psychology, inhibition is located in the individual and violence is a result of the *breakdown of intrapsychic inhibition*. This breakdown is attributed to a depletion of individual resources that weaken personal control. In sociology, inhibition is located in the interaction and violence results from the *breakdown of intersubjective inhibition*. This breakdown is attributed to the particular weakness of the victim in an interaction or the presence of an audience who undermine normal intersubjective control mechanisms. In primatology, inhibition is located in the evolutionary importance of social relationships and violence results from the *breakdown of valued relationship inhibition*. This breakdown is attributed to a calculation of the benefits of violence over the costs of violence for individual gain. This is offset by a natural conflict resolution mechanism that institutes repair.

Having compared the foundations that underpin the transition to violence in different disciplines, we now turn to more specific research on the ability of third-parties to shape the trajectory of violence. In doing so, we move from a focus on where to locate the source of inhibition of violence to an exploration of the evidence involving the direct impact of third-parties on the likelihood of violence. As before, we will consider evidence from the fields of psychology, sociology and primatology.

1.5.1 Psychology and Bystander Behaviour

In psychology, the presence of others has traditionally been seen to have a negative impact on the behaviour of individuals. For example, deindividuation theory (Diener, 1976; Zimbardo, 1969) suggests that perceptions of anonymity related to the co-presence of others leads to a lack of accountability that in turn creates the conditions for loss of selfhood and behavioural control. This loss of control is likely to result in more antisocial behaviour and violence. At the same time, the presence of others is also argued to inhibit prosocial behaviour. For example, the bystander effect literature (Latané & Darley, 1970) suggests that the presence of others leads to diffusion of responsibility, pluralistic ignorance and audience inhibition. The argument here is that individuals on their own are likely to intervene to help, but are less likely to do so in the presence of others—an effect that increases with group size. Thus, it would seem that the presence of others during an aggressive event creates the perfect storm of promoting violence and undermining willingness to intervene.

However, in more recent times, researchers—particularly those influenced by the social identity approach (Tajfel & Turner, 1979, 1986; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987)—have provided evidence that challenges these preconceptions. In the mid-1990s, Reicher, Spears, and Postmes (1995) offered their social identity model of deindividuation effects (SIDE) as an alternative to deindividuation theory. Unlike traditional deindividuation theory (Festinger, Pepitone, & Newcomb, 1952), which stipulates that group members become liberated from social responsibility when submerged collectively, Reicher et al. (Reicher, 1996; 1995) suggested that deindividuation frequently resulted in increased responsibility and normative behaviour. This was empirically supported by the work of Postmes and Spears (1998) who conducted a meta-analysis of 60 studies and concluded that the effects of deindividuation played a critical role in creating combined social regulations and collective behaviour, not social deterioration. Put simply, it is the norms of the local identities that shape when anti-social behaviour may occur. The presence of others does not

lead to a loss of identity (and thus loss of normative control) that in turn results in violence. Instead, if violence occurs, it emerges through situationally-relevant social identities that govern which behaviours are acceptable and which targets are legitimate (Drury & Reicher, 1999; Reicher, 1984).

In a similar fashion, a meta-analysis by Fischer et al. (2011) showed that the bystander effect does not hold for emergency situations that are dangerous or violent. Fischer et al. (2011) suggested that this might be a consequence of the greater arousal that attends these kinds of incidents. However, Levine et al. (2010; Levine et al., 2011; Levine & Crowther, 2008) proposed an alternative social identity-based argument, which suggested that it is the local identity-based norms, combined with the collective strength of the group, that explains the absence of the bystander effect. Levine et al. (2012) further postulated that in interpersonal aggressive encounters attended by third-parties, social identities become salient, as too do the norms and values of those identities. These norms shape the potential likelihood of bystanders intervening to prevent violence in the same way that they might escalate violence. Specifically, in Levine and colleagues' (2012) interview work with patrons of the NTE, interviewees frequently described situations in which interpersonal conflict and violence intervention were shaped through the salience of common membership with ingroup members (e.g. 'we're all here having a good night together, let's not fight') as well as by intergroup comparisons with outgroup 'others' ('we can't just stand back when our mate's in trouble with them'). Indeed, Levine et al. (2011) showed in these NTE contexts, using a behavioural microanalysis of CCTV footage of street violence, that increased group size results in more de-escalation, rather than escalation of aggression, and that successful violence prevention is better predicted by the cumulative action of bystanders than by a focus on the perpetrator or victim alone.

1.5.2 Sociology and Guardianship

In sociology, or, more specifically, the microsociology of Collins (2008), we have already seen that the presence of an audience is one of the critical features that can allow confrontational tension and fear to be overcome, resulting in violence. This approach to the presence of others tends to emphasise the negative, rather than the positive, impact of third-parties on violence (akin to the traditional deindividuation theory in psychology). However, there is also a longstanding tradition of scholarship in sociology that emphasises the importance of ‘guardianship’ in the control of behaviour and prevention of crime. Routine activities theory (Cohen & Felson, 1979) posits that there are three essential components—which can co-occur during typical routine activities in everyday life, and which can provide the conditions for crimes to emerge: a motivated offender, suitable targets, and the lack of capable guardianship. This triadic constellation at the centre of crime gives the role of third-parties equal weight with that of perpetrators and victims. Over the last 30 years, there has been much work to define the concept of ‘guardianship’, and it has now come to be defined as “the presence of a human element which acts—whether intentionally or not—to deter the would-be offender from committing a crime against an available target” (Hollis, Felson, & Welsh, 2013, p. 76). From this perspective, the role of third-parties in violence seems to favour social control and regulation, rather than disorder and escalation. More recent work on guardianship in action (GIA; Reynald, 2017) has made the case for the importance of understanding the everyday dynamics of the role of guardians on the likely occurrence of crime. Therefore, in contrast to the social psychological literature, this theoretical tradition advocates that social inhibition stemming from the presence and actions of ‘others’ are in fact critical for maintaining social order, with delinquency arising as the result of a failure or breakdown in this social control (Janowitz, 1975; Reiss, 1951). The same argument can be made for understanding the human transition to violence.

1.5.3 Third-parties in Non-Human Primate Communities

In the non-human primate literature, recent developments on ‘triadic social awareness’ (de Waal, 1982) have demonstrated that social animals involved in conflict can orientate not only towards antagonistic others, but also towards third-parties (Fedurek, Slocombe, Hartel, & Zuberbühler, 2015; Wittig, Crockford, Langergraber, & Zuberbühler, 2014). For example, research on vocalisations during conflict showed that primates directed their screams towards conflict partners (as pleas or deterrence) and towards the audience (as appeals for support) (Fedurek et al., 2015; Slocombe & Zuberbühler, 2007). Recent research by Flack et al. (Flack, de Waal, & Krakauer, 2005b; Flack, Girvan, de Waal, & Krakauer, 2006; Krakauer, Page, & Flack, 2011) on ‘third[-]party policing’ further examined how bystander animals intervene in conflicts to achieve collective third-party conflict resolution. More specifically, this body of research showed that when aggression broke out between two animals in a troop, a third animal (or more) entered the engagement and policed those who were agitated back into a passive state. This community-based regulation is found to have varying degrees of success depending on the hierarchical structure of the society, with asymmetrical alpha-dominated societies having greater policing success rates than more democratic animal groups (where there is less agreement regarding the main sources of authority). In this work, Flack et al. (Flack et al., 2005b; Flack, de Waal, & Krakauer, 2005a; Flack et al., 2006; Krakauer et al., 2011) argued that the existence of third-party policing—intervening impartially to control conflict—is something of an evolutionary puzzle. Policers pay a cost through their interventions, while the benefits of such an intervention are distributed across the group. More recently, Flack and colleagues’ (2005b, 2005a, 2006) data on third-party policing in pig-tailed macaques have been used to develop game theoretic models of the dynamics of conflict. For example, DeDeo, Krakauer and Flack (2010) used a Monte Carlo simulation technique to test alternative causal models of conflict dynamics. They compared models in which the primary driver was a ‘rogue individual’, a pair of interactants or a triadic relationship between the

third-party and the antagonists. They found that the best fit to their macaque conflict data was a triadic solution, whereby the social assessments, on which decisions to fight were based, were not determined by individual agency, nor by pairwise calculations. The authors then went on to suggest that the fact that “the causal unit of conflict dynamics is the triad, not the individual nor the pair, suggests that individual agency has been overemphasized in social evolution” (p. 13). Put simply, in social species, social evaluations in violence should matter.

It is important to note, however, that the above conflict resolution ethology literature exclusively depicts intragroup encounters, largely drawn from captive populations. In contrast, the non-human primate *intergroup* literature reports a notable lack of empathic responses shown towards outgroup conspecifics (M. W. Campbell & de Waal, 2014) and a complete absence—both experimentally and in the field—of intergroup peacekeeping behaviours. Rather when two foreign groups meet they do so with great suspicion, orientating protectively towards their own and aggressively towards others (Muller, 2002; M. L. Wilson & Wrangham, 2003). Triadic evaluation in intergroup conflict appears mostly restricted to assessments of numerical advantage over an opponent that reduces the relative risk of serious injury (Sakamaki & Nakamura, 2015).

1.6 Social Group Categorisation and Its Effect on Aggressive Interactions

As we can see from this review of the psychology, sociology and primatology literatures, explanations for the transition to violence need to incorporate actors beyond the individual and the dyadic relationship. All these disciplines suggest that, at the very least, an acknowledgement of the importance of interactions at the level of the triad is key to conceptualising aggression to violence as a dynamic process.

If it is acknowledged that aggressive interactions are embedded within a local social context, then it is also important to recognise that social relationship ties (Swann, Jetten, Gómez, Whitehouse, & Bastian, 2012) and salient social group memberships (Turner et al., 1987) exist between the relevant actors (protagonist, target and bystanders) within a conflict.

Social relationship appraisals may influence how individuals interact with one another as a conflict unfolds. This is identified by Otten (2009):

Knowing which group a person belongs to, and whether he or she 'is one of us' or 'one of them', provides us with heuristics on how to give meaning to certain events, and on what to expect in further interaction. This should definitely also apply to the way we perceive and react in aggressive interactions. (p. 162)

Before we dissect this claim, it is important to clarify how a social group is defined and how social group categorisation may impact on cognitions and subsequent behaviour.

While it is true that as individuals we all possess an idiosyncratic identity, when collected together with others we also categorise ourselves as members of particular social groups (self-categorization theory; Turner et al., 1987) that cognitively become part of the self (E. R. Smith & Henry, 1996). It is theoretically assumed that social groups are represented mentally as prototypes (Hogg & Reid, 2006). These prototypes are fuzzy sets of interrelated attributes (attitudes, feelings and perceptions) that capture the similarity within a group and accentuate the differences with those not within the group, thus discriminating the relevant category from outgroups (Hogg, 2006). As such, a prototype sets the boundaries surrounding an identity and determines who may be considered ingroup or outgroup. Beyond setting the inclusion boundaries of a social group (who is ingroup and who is outgroup), a prototype also describes and prescribes how a member ought to behave (Hogg & Reid, 2006). "In this sense prototypes are norms; that is, because a particular perception, behaviour or attitude is shared within a group, it is normative of that particular group [...] thus, prototype-based attitudes are normative" (J. R. Smith & Hogg, 2008, p. 344). During categorisation and identification with a group containing the self (ingroup), and through a process referred to as depersonalization, individuals typically subsume these prototypically-relevant attributes (Turner et al., 1987). This assimilation leads to individual behaviour congruous with that of the group, with an adherence to collective norms, values and expectations.

Through categorisation and identification with a group containing the self (ingroup), social comparison with other groups (outgroups) becomes inevitable (Turner et al., 1987). This mere categorisation of ‘us’ and ‘them’ has an impact on behaviour, affect, motivations and cognition (for a review, see Dovidio & Gaertner, 2010). Specifically, since individuals seek a positive assessment of their idiosyncratic self and, by extension, their collective self, group members have a tendency to evaluate their own ingroup and its members more positively than the relevant outgroup (Brewer, 1979; Otten & Moskowitz, 2000) and to trust their own members over others (Insko et al., 2001; Kramer & Cook, 2004). The need for a positive self-evaluation, together with social comparisons that favour one’s ingroup, is further associated with ingroup preferences, unfair hoarding of resources and a motivational basis for supporting ingroup members’ efforts over others (Brewer, 2002; Hewstone, Rubin, & Willis, 2002; Tajfel, Billig, Bundy, & Flament, 1971).

Returning to the domain of aggression and violence, the question remains whether the categorisation of ‘us’ and ‘them’ has sufficient impact on cognitions, motivations and subsequent behaviours related to intergroup hostility and intragroup support. At first appraisal, this may seem easy to answer: in a between-group conflict where we categorise ‘us’ and ‘them’, it is essentially ‘us’ versus ‘them’. Social categorisation may result in a *positive ingroup bias*, whereby the negative behaviour of an ingroup member is perceived as comparatively less hostile, less intentional and more easily justifiable than the same behaviour of an outgroup member (Brown & Capozza, 2006; Otten, 2009). Otten and Jacoby (2004) found experimental evidence for this bias in a hypothetical scenario in which Dutch participants imagined a Dutch football fan being tripped over by either a German or fellow Dutch supporter. Although the intentionality of this act was kept ambiguous, when the offender was an outgroup member, Dutch participants perceived the trip as more intentional and more hostile than when the offender was an ingroup member. Such findings have since been replicated in a hypothetical scenario of a painful shove, either enacted by a person from

one's own school (ingroup member) or by a person from another school (outgroup member) (Knuijver, Kornfeld, Ravensbergen, Timmer, & Witteman, 2006).

Beyond positive valence towards one's own intentions, there is also ample empirical work to suggest that individuals have a predisposition to feel concern for, and to support, their own group members by default over others (for reviews, see Reicher & Haslam, 2009; Stürmer & Snyder, 2009), particularly during emergencies (Levine & Manning, 2013). In the context of violence, Slater et al. (2013) demonstrated that shared social group membership with a victim increased the likelihood of bystander intervention in an immersive barroom fight scenario. Most recently, social group membership has been identified as an important predictor of peaceful bystander interventions during real-life street violence in Copenhagen (Liebst, 2017 in preparation) and in the aftermath of violent robberies in the Netherlands (Lindegaard et al., 2017). In this series of papers, social group membership was markedly the strongest predictor of intervention, out-predicting more conventional variables of gender, number of bystanders and severity of situation (Fischer et al., 2011).

However, not all interventions on behalf of group members are peaceful. Group members can be supportive audiences of intergroup conflict, who passively 'allow' violence to occur by non-intervention (Levine, 1999). Group members can provide verbal support to their fellow group members in combat and encourage the protagonists to engage (Collins, 2008; Felson, 1982). Group members can also be partisan fighters who join in an ongoing conflict for collective thrills, or to protect their own group members from harm (Parks, Osgood, Felson, Wells, & Graham, 2013; Weenink, 2014). In gang violence research, increased perceived attachment to ingroup members under attack has also been found to prompt greater social unity (Decker, 1996), heightened perceptions of outgroup entitativity, and thus a higher willingness to retaliate to threats with violence (Vasquez, Wenborne, Peers, Alleyne, & Ellis, 2015). Further, an attack on a fellow ingroup member can be perceived as an attack on the entire group's reputation and honour (Tedeschi & Felson, 1994; Vasquez et al.,

2015). As such, and in order to maintain positive group identity, groups may retaliate against outgroup others ‘for the sake of the group’, which can function to save face, even the score and gain dominance. It is theorised that the likelihood of partisan hostility becomes most likely when the social distance to the group member in combat is close, and the social distance from the intended target is far (Black, 2014). This accumulated evidence would suggest that under certain conditions, socially categorised third-parties may be more likely to contribute to intergroup conflict escalation than pacification.

However, assumption of an automatic negative response to outgroup others in conflict, and an automatic supporting role to ingroup others, may be a gross simplification. An integral facet of the social identity approach is that identity encompasses the norms and values that establish what is meaningful and appropriate for the group members (Reicher & Hopkins, 2001). When group membership becomes salient, members will comply with the content of this specific identity and follow the norms of that particular group (Livingstone & Haslam, 2008). As such, members ask: “what is appropriate for us as members of this category in this context?” (Reicher, 2001, p. 195). If violence is something that goes against the grain of the salient identity, then it is behaviour unlikely to be approved by other group members, or utilised.

In the experimental work of Otten and Jacoby (2004) and Knuijver et al. (2006), a trip or a physical shove were perceived as less intentional and less harmful when originating from an ingroup as opposed to an outgroup member (*positive bias*); however, these effects were found to reverse once information was provided that the fellow group member had *intended* to harm the other. When an ingroup member intended to purposely harm another individual, this was interpreted as a violation against the positive expectations of the group, and thus resulted in a pronounced negative reaction (black sheep effect, see Marques, Abrams, Páez, & Hogg, 2001). In such a scenario, “the positive valence automatically associated with one’s own

group [...] is exceeded by the negative valence associated with the critical behaviour” (Otten, 2009, p. 177).

In the same vein, there is accumulating evidence outside of experimental settings of groups ‘self-policing’ their own deviant group members, whose violent actions go against the norms and values of the group (Reicher, 1987; Stott, Adang, Livingstone, & Schreiber, 2007; Stott, Hutchison, & Drury, 2001). For example, during the 1998 World Cup, two Scottish football fans were dragged away from their violent assault of a Tunisian fan by fellow Scots and punished accordingly (Stott et al., 2001). When asked in an interview why they had intervened, the fellow Scots suggested that the deviants had threatened the content of the ‘Scottish identity’. More specifically, by being violent towards the Tunisian, the Scottish deviants were acting in line with ‘racist hooligan’ English fans. This position was far removed from the fun-loving and friendly image the Scots liked to project and deemed unacceptable. Likewise, interviews with drinking patrons of public spaces found that group members would reportedly ‘police our own’ during aggressive episodes to deter deviants from acting violently (Levine et al., 2012). The array of intervention techniques was vast and ranged from ignoring or babysitting the misbehaving individual to physically restraining or even attacking him or her. Discussions of these interventions were underpinned by implicit rules, and bystanders and group members knew which acts of a perpetrator were legitimate and which warranted intervention (Levine et al., 2012). Even the more extreme examples of inflicting harm to one’s own group members were imbued with implicit rules and social legitimacy. In these examples, violence was used for the sake of nonviolence, and harm was inflicted as a physical communication of collective disapproval and, as such, retributive justice for group norm violation (see Vidmar, 2001).

Even when group violence does occur between groups, it is not mindless. Rather, the violence is governed by group values that determine which behaviours are acceptable and which targets are legitimate. For example, during the 1980 St. Paul’s riots, it was considered

acceptable when an ingroup rioter threw a stone at the police (a figure of opposing authority) and the action was repeated by other category members. However, when another ingroup rioter threw a stone at a bus (an illegitimate target), others failed to join in (Reicher, 1984). In accordance, during the 1990 anti-poll tax demonstrations in Exeter, outgroup members failed in their attempt to incite violence within the general crowd, who collectively and within their subcategories had distinguished which behaviours were deemed appropriate (Drury & Reicher, 1999).

Together, such findings stress the contribution of group processes to the understanding of emergent violence and its regulation. Whether the actions of third-parties produce violence or peace, in the first instance, seems to be qualified by whether the aggression happens in intergroup or intragroup contexts. However, this relationship is multifaceted, and actions from individuals are likely to be contingent on perceived normative approval and socially-bound legitimacy. Therefore, to understand when violence will occur, researchers need to look beyond the dyadic pair and understand the relative contributions and interactions of all relevant actors in context.

In summary, several bodies of work on third-parties in violence have made strong theoretical cases for the need to situate violence within its immediate situation and local social context. These alternative strands of research advocate that to understand violence in public spaces, it is inadequate to only consider the dispositional components or dyadic interaction of the protagonist and the target, as models of aggression chiefly do. Rather, researchers need to look beyond the dyadic pair and understand the relative contributions and interactions of all actors and the importance of third-parties to violence.

The current state of the literature provides two main overarching questions. First, while mainstream models of aggression and violence have largely ignored the social context as a predictor of violence, alternative strands of research suggest it is imperative to look beyond the aggressive dyad: can this be examined? Second, while there are some theoretical

assumptions that social categorisation (i.e., the extent to which aggression involves intragroup or intergroup dynamics) should impact on escalatory/de-escalatory behaviours and conflict outcomes, do social categorisations really matter? These two questions make up the central construct of this thesis.

However, methodologically and ethically, research into violence is extremely challenging (as will be discussed in the next chapter). Research outside of the laboratory into the role of social groups and third-parties in public space violence is not exempt from these challenges. Therefore, the next question to address is how best to study violence in public spaces. Further, how have the findings from the empirical methods available contributed to our understanding of the role of social groups and third-parties as important explanatory components of public space violence?

2 Analysing Violence in the NTE: A methodological review and systematic behavioural analysis of intergroup-intragroup violent interactions

The aims of this chapter are to introduce CCTV visual data analysis as a viable and constructive methodological tool for the study of human violence, and to employ this method to empirically examine the importance of social groups and third-parties as explanatory components of public space violence.

We first review the different methodological approaches applied to research on interpersonal violence in the night-time economy (NTE), with the dual purpose of evaluating the strengths and limitations of the field's established methodologies and proposing a research agenda based on the microanalysis of CCTV footage. We then conduct our first empirical CCTV study to ascertain whether a de-escalatory bias towards one's own group members and an escalatory bias towards outgroup members exist in episodes of public space violence. Furthermore, we examine the level of involvement of third-party bystanders, and the rates of their escalatory and de-escalatory contributions. Notable effects of the 'group' on escalation and de-escalation provide further support for the need to consider social group membership as a unit of analysis in models of aggression and violence in public spaces. Moreover, notable influences of third-parties emphasise the necessity of looking beyond the aggressive dyad (protagonist and target) to examine how third-parties may affect violent outcomes.

The rise of the NTE is associated with a rise in public space violence and subsequent public anxiety (Bromley et al., 2000; Hadfield et al., 2009; Lister, 2009). Research aimed at better understanding and thus reducing this violence has employed a range of empirical methods, such as official records, victim and patron interviews, laboratory aggression studies and direct observations. Although informative, none of these methods reliably capture the detailed interactions that occur between perpetrators, victims and bystanders of violence. In this methodological review, we identify the key methodologies employed to examine violence in NTE contexts. We evaluate the strengths and limitations of each method with regard to

analysing violent interactions. We suggest that CCTV behavioural analysis is a promising, yet under-utilised, method for studying violent behaviour in the NTE. Specifically, violence research that has analysed video footage has found interactions to be important situational factors for understanding enactments of violence that involve more than two parties. Detailed microanalysis of real violent interactions has enabled researchers to challenge the orthodoxy found in the social psychological literature, where group processes are seen to create the conditions for violence to emerge. Simultaneously, this method has allowed us to explore the patterns of social regulation of violence predicted by sociological theory, which have hardly been studied directly by sociologists. These interactions have not been captured by traditional research methods. We conclude by considering the limitations and ethical considerations of using CCTV footage for research.

2.1 Established and Emerging Methodological Approaches

2.1.1 Official Records

The principle source of data regarding violence in the NTE is police-recorded crime. This data offers a wealth of information on levels of violence, changes over time and the geographical hotspots in which injuries are most prevalent (Burrell & Erol, 2009; Hadfield et al., 2009). For example, Newton and Hirschfield (2009) examined the relationship between alcohol trading hours and occurrences of violence and found that the analysis was hindered by the lack of consistency in the format of the collected data. In later research, Newton (2015) found a strong spatial correlation between licensed premises and registered violent crime. Loveday (2005) found an increase in ‘stranger violence’ following the extended opening hours permitted by the introduction of the *Licensing Act 2003*.

However, this type of data suffers from a selection bias, because only the incidents witnessed on police patrols, or directly reported to the authorities, are recorded. The likelihood of incidents in the NTE being self-reported to authorities may be inhibited by

intoxication. Brennan (2011) examined the individual, situational and alcohol-related risk factors associated with reporting violent victimisation in the NTE. He noted that some drinkers engage in a 'moratorium' on reporting violence when socialising and intoxicated, leading to fewer reported incidents. The final sample of officially-registered offences is therefore likely to be restricted to high-severity incidents and those cases deemed most court-worthy. These limitations aside, official records do not provide details on individuals' own constructions of violence, or of how people behave as an aggressive episode unfolds. To better understand the incidents that are more likely to lead to violence, researchers have analysed the motivations of the people involved in aggressive episodes through qualitative interviews.

2.1.2 Interviews

Interviews and verbal accounts of violence in the NTE complement official statistics by offering a greater description of the real experiences and behaviours enacted during aggressive episodes (Levine et al., 2012). For example, Graham and Wells (2003) interviewed 21 young males and proposed a typology of fighters: the self-proclaimed non-fighter that prefers vocal aggression, the necessary evil fighter, the grievance fighter and the recreational fighter. These 21 male interviewees largely considered fights acceptable and their participatory actions normative. Further, fight participation provided individuals with rewards in the form of attention from third-parties, increased group cohesion and euphoric rushes from feeling heroic. Lowe et al. (2012) conducted group interviews exploring the presentational concerns of both male and female bystanders when intervening in public displays of female-to-female violence. Male bystanders had a tendency to laugh at female-to-female violence, and expressed concerns over inflicting unintended injuries or of receiving accusations of sexual harassment if they were to intervene. Females reported a greater motivation to intervene than males; however, this motivation derived from a concern that female fighters would shame or sexualise *all* women. Copes, Hochstetler and Forsyth (2013) analysed 23 interviews with males who had been in several bar fights and identified four common

purposes of fighting: verifying a self-image of masculinity, deterring potential assailants, maintaining hierarchies and settling matters. These interviewees specified rules of conduct: do not seriously harm the other fighter, fight others of similar physical size and fight in equal numbers on both sides. Abiding by these guidelines was a way of constructing one's identity as a person deserving of respect.

The limitations of interview studies are well known and are accentuated by circumstances of the NTE setting. Participant response bias implies that more socially-desirable answers are expressed and negative behaviours are understated. This is especially pertinent for sensitive and socially disapproved topics including aggression and violence (Gregoski, Malone, & Richardson, 2005; Saunders, 1991). Further, retrospective recalling and self-reported accounts of violent episodes may be inaccurate owing to cognitive constraints, false memories, self-deception and memory failure—phenomena that have been found to increase in times of high stress, aggression and alcohol intoxication (Laney & Takarangi, 2013; Saunders, 1991; Vrij, Hope, & Fisher, 2014; Yuille & Tollestrup, 1990). Additionally, as violence in the NTE often serves as a symbolic means of status and masculinity validation, interviewees may provide dramatised narratives and distorted self-presentational accounts of actual events (Collins, 2008; Hochstetler, Copes, & Forsyth, 2014; Jackson-Jacobs, 2004).

2.1.3 Laboratory Settings

To circumvent these complications and to identify the causal factors influencing violence in the NTE, scholars have advocated the study of aggression paradigms in controlled laboratory settings (Anderson & Bushman, 1997). These studies involve experiments that provoke a participant under varying conditions and measure the subsequent aggressive behavioural responses. An impressive list of personal (e.g., gender, individual beliefs, attitudes, genetic predispositions, goals and personality traits) and situational (e.g., pain and frustrations, heat, aggressive cues and provocations) risk factors have been identified from

this research (Slotter & Finkel, 2011). The aim is to understand the underlying mechanisms of aggressive behavioural responses so as to apply it to all types of violence and aggression.

Experiments typically measure aggression using dependent variables such as hot sauce doses, mild shock or noise blast administrations, written retaliatory insults, negative feedback scores and even virtual pins in virtual voodoo dolls (see Finkel, 2014). Although novel approaches are required owing to ethical restrictions, it remains debatable whether these responses (which may be inadvertently cued by the experimenter; cf. Ritter & Eslea, 2005) are in fact aggressive, violent or perhaps even amusing. The results from these experimental studies have been applied to violence research across multiple domains, from intimate partner violence to racial genocides and suicide to torture (Benjamin, 2008; DeWall et al., 2011). However, these risk factors have been found to be less effective in describing or predicting real-world aggression (Ferguson & Dyck, 2012).

Recently, scholars have attempted to enhance the real-world validity of laboratory studies with the use of virtual reality technologies (van Gelder, Otte, & Luciano, 2014). For example, Slater et al. (2013) found evidence that social group membership may increase the likelihood of bystander intervention in an immersive bar-room fight scenario. Putting aside recent virtual reality advancements, a ‘one-size-fits-all’ model of violence that is assumed across these laboratory studies inevitably reduces “the complexity, diversity, and contextual aspects of real-world aggression” (Graham et al., 2006, p. 281).

2.1.4 Naturalistic On-Site Observations

In light of these shortcomings, Archer and Browne (1989), and more recently Graham et al. (2006), have called for research into violence to include naturalistic observation. A trained observer can witness and discreetly record (often in notepads, on coding sheets and via voice memo recordings, etc.) the actions of those involved in the aggressive episode as it unfolds. The advantages of this method are apparent. Unlike interview data and verbal accounts, naturalistic observation circumvents concerns associated with participant response

bias and demand characteristics. In contrast to laboratory experimentation, naturalistic observation captures real-life aggression in its contextual setting, thus providing greater ecological validity. Further, similar to experimental work, naturalistic observations permit the researcher to record quantitative counts of certain behaviours, thus allowing for quantitative analysis. Finally, contrary to police report data—which tend to only record the most court-worthy, violent incidents—naturalistic observations can capture conflicts that vary in severity, from mild altercations and posturing that dissipates naturally through to incidents that escalate into more serious outcomes. This means that the researcher is not restricted to severe cases that have little to no variation on the dependent variable.

Naturalistic observation, applied either as a quantitative-systematic or qualitative-ethnographic technique, has allowed researchers to better understand the environmental, situational and social features of the NTE that contribute to conflict between patrons. For example, systematic observations have identified relationships between environmental features—such as spatial design, cleanliness, loitering and people density (e.g., crowding, peak times and cueing)—and levels of aggression (Graham et al., 2004; Graham & Homel, 2012; Macintyre & Homel, 1997; Townsley & Grimshaw, 2013).

On-site observations have highlighted how bar management styles and serving practices can be important predictors of violence (Graham et al., 2004; Homel, Clark, & others, 1994; Homel & Tomsen, 1993). This research has also demonstrated how social features such as masculinity and gender can explain differences in displays of aggression; masculinity encourages men to accept challenges from other men (Forsyth & Lennox, 2010; Graham et al., 2006; Graham & Homel, 2012; Tomsen, 1997). Finally, research using naturalistic on-site observations has provided insights into situational features (e.g., the severity of aggression) that lead bystanders to intervene in bar-room aggression (Parks et al., 2013). It should also be added that naturalistic on-site observation has also been applied within ethology to examine similar peacekeeping behaviours in non-human primates. In a

pioneering study of chimpanzees, de Waal and van Roosmalen (1979) observed patterns of friendly interactions between former opponents, as well as friendly interactions of third-party bystanders who assisted in easing conflict distress by hugging and touching the victim of aggression.

Despite the empirical advances owing to this method, naturalistic observations are limited through their dependence on the sightings of a single or often pair of on-site coders. Human coders are never a *tabula rasa*, but instead enter any observational setting with preconceived stereotypes, biases and beliefs. Coders have certain expectations of the personal characteristics and behavioural tendencies of those they encounter and may observe according to these preconceptions. Confirmation bias may occur in the data, as human coders unconsciously look for, and interpret, information that is consistent with their research question, and discard or ignore evidence that does not fit with their hypothesis. The coders—who often spend several hours observing on site—may also experience effects of fatigue and lapses in attention, which, again, result in lost information.

Further, coders cannot observe from within a bubble; rather, they are intermingled with the immediate situation surrounding them. One benefit of this is that it allows insight into the experiences of the people they are observing. However, the presence of an observer may inhibit or disturb the natural behaviours of those that the observer intends to code (LeCompte & Goetz, 1982). This may be particularly problematic in the study of public aggression, whereby the severity of violence may alter as a function of increased numbers and the presence of an audience (Felson, 1982; Levine et al., 2011). To remain as inconspicuous as possible, coders may opt to abstain from making detailed notes until after their observation has finished. However, this delay exposes observations to issues of false memories, unconscious self-deceptions and memory failures that are usually attributed to self-report methods. Owing to the complexity and chaotic nature of violent episodes, and the large number of people simultaneously contributing to the outcome (Levine et al., 2011; Nassauer

& Legewie, 2016), it is impossible for the human observer to account for (particularly in a chronological or second-by-second order) all behaviours exhibited by all actors. This makes it difficult to construct detailed accounts of all the actions that unfold. Even the deployment of multiple on-site coders cannot address these concerns or ensure entirely reliable observations (see discussion of Granstrom et al., 2009, in Adang, 2016), which means that the details of what is occurring between those in and around the situation may be lost.

2.1.5 Video Analysis as an Observational Method

A way to circumvent the over-reliance on a coder in situ, and yet retain the numerous benefits of observational analysis, is to observe natural behaviours through prerecorded videos. In recent years, there has been a vast increase in CCTV surveillance, with a conservative estimate of at least 1.85 million CCTV cameras in Britain today—one in operation for every 34 people (Gerrard & Thompson, 2011). While this increased camera presence has assisted in identifying and prosecuting violent offenders, and helped to reduce crimes such as vehicle break-ins and shoplifting, the British Home Office's own reports have found limited evidence to suggest that having this camera presence deters violent crime (Welsh & Farrington, 2002, 2009). However, the footage from these video-recording devices offers vital insight into how attackers and victims use public spaces, and is a valuable resource for recording and subsequently analysing real aggressive episodes in humans.

A well-placed CCTV camera does not require the presence of a human observer and has a radius of sight and line of focus greater than a human observer. A camera can also indiscriminately record all unobscured actions within its field of view, without fatigue or lapses in concentration. Video data can also be replayed, rewound, paused or slowed down to frame-by-frame instances. This provides a second-by-second account of the events that unfold and allows iterative identification and inclusion of behaviours that may have been missed if only viewed once. Further, such observational control enables checks in inter-rater reliability between coders and the verification of existing coding schemes, enabling a rigorous and

highly detailed systematic analysis—a necessity long advocated in the observational tradition (Johnson & Sackett, 1998). Video analysis allows the researcher to exploit the chronological chain and temporal properties of the interaction sequence. This means that video analysis can accommodate specific statistics that otherwise could not be utilised, including sequential and survival analyses (examples of such are demonstrated in the discussion of Levine et al., 2011). This allows researchers to address and challenge many of our preconceptions of violence, including ideas around how groups of people and bystanders shape aggression.

2.2 Group Behaviour in NTE Violence—New Insights and Avenues for Future

Research Using CCTV

Traditionally, in social psychology, groups have been seen to create the conditions for violence as a result of the irrationality, anonymity and lack of accountability they can engender in individuals (e.g., Diener, 1976; Le Bon, 1897/1960). Moreover, research into the bystander effect (Latané & Darley, 1970; for a comprehensive review, see Fischer et al. 2011) suggests that increases in group size result in a greater diffusion of responsibility and a decreased likelihood of prosocial intervention (Colman, 1991; Littman & Paluck, 2015; Zimbardo, 2008).

These ideas have contributed to the public opinion that Britain is inundated with uncontrollable groups of people, who go out drinking and fighting at night (Crawford & Flint, 2009; Roberts, 2006). This has created a social and moral anxiety that has shaped recent policing practices that encourage the dispersion of large groups in public spaces and at sporting events, and drove the call for more police on the beat (Porter, 2015). However, if authority is needed to control groups and prevent violence in public spaces, then the ratio of formal guardian to patron is inadequate (Hobbs, 2003). As an example, in the British city of Manchester, depending on the state of the economy, there are an estimated 130,000 people in NTE leisure areas circuit drinking on weekend nights (Hobbs et al., 2005). In total, there are approximately 40 on-duty police officers and around 1000 bouncers, who are often

undertrained and impassive to the actions that take place outside of their premises (Levine et al., 2012). The current state of the social psychological literature predicts that the excitement, anonymity and reduced responsibility afforded by increased numbers—coupled with reduced inhibition from alcohol and other potential drug substances—will result in high levels of aggression and violence. With this in mind, we would expect chaos to reign in our public spaces every single weekend, but the reality is different. Of course, it would be naive to suggest that fighters and victims of violence are not occupying emergency services each weekend and during many national holidays (see Bellis et al., 2012), but what is interesting to note is how little violence there is given the large numbers of patrons (Marsh & Fox-Kibby, 1992). With limited formal authority figures in these public spaces, and with the CCTV coverage seemingly ineffective at reducing violent crime (Gill & Spriggs, 2005; Welsh & Farrington, 2002, 2009), there must be another factor that is ‘policing’ and keeping control of groups and preventing violence in public spaces.

Insights from sociology may provide answers to this question. Contrary to the viewpoint that groups incite violence, a body of sociological research on ‘social control’ suggests that ‘others’ are in fact critical for creating the conditions necessary for social regulation (Goff, 2014; Janowitz, 1975; Reiss, 1951). ‘Others’ may help curtail deviant behaviour through the direct surveillance and supervision of potential deviants, and through actions and interventions that signal that certain behaviours are unacceptable (Goff, 2014). In contrast to the social psychological literature, this theoretical tradition advocates that delinquency arises as the result of a failure or breakdown in this social control (Janowitz, 1975; Reiss, 1951). However, although there are assumptions that social groups may enforce social control (for the counterargument that third-parties may instead reinforce dispute-related violence, see Felson, 1993), these expectations have not translated into a direct systematic investigation of the role of others in violent events.

Further, current sociological research—which looks more broadly at the associations between reported levels of social control in geographical locations (e.g., in neighbourhoods, communities and street segments; Groff, 2014; Weisburd, Groff, & Yang, 2012) and increased delinquent behaviour—has not yet demonstrated exactly how others who are present coordinate their actions to ensure that such informal social control is obtained. For microsociologists, it is these specific interaction rituals and microbehaviours between actors in the immediate situation that provide the greatest insight into social order maintenance and the escalation of violence (Collins, 2008).

Contrary to the negative group orthodoxy present in social psychology, and in line with the descriptive assumptions of sociology's social control, Levine et al. (2011) hypothesised that not only does group regulation take place in aggressive episodes in public spaces, but that bystanders may be important players who can de-escalate the conflict and shape the trajectory of violence. Levine et al. (2011) conducted a systematic behavioural analysis of 42 CCTV clips of public space aggression that did not involve the intervention of an official authority figure (e.g., police officer, door security staff or community warden). This dataset provided access to the naturalistic behaviours of those involved in, and bystanders of, real aggressive episodes in public spaces. Further, this dataset followed strict exclusion criteria to ensure that each clip captured aggressive events in their entirety, from the point at which the CCTV camera operative first noted hostility through to the conclusion of natural conflict dissipation or police presence.

Using the video software Observer XT (Noldus, 1991), an observational analysis tool used in the social and biological sciences for behaviourally coding video data, Levine et al. (2011) coded the observable actions of all those present within the 42 clips, which lasted between one and eight minutes. Individuals were classified either as the *protagonists* driving the initial aggression, *targets* receiving the aggression or *third-party bystanders* who interacted at least once with anyone else in the clip. Actions were divided into *physical*

escalatory acts, non-physical escalatory acts or de-escalatory acts. Physical escalatory acts included kicking, punching, slapping, grappling and pushing (see Figure 2.1). Non-physical escalatory acts included aggressive gesturing, pointing, invading personal space and removing one's own clothes. These two categories (physical escalatory and non-physical escalatory) were then aggregated before analysis into a single category: 'escalatory acts'. De-escalatory acts included calming open-hand gestures, movements to block contact, pulling people apart and holding people back (see Figure 2.1).



Figure 2.1. Examples of an escalatory behaviour in the form of a punch (left) and a de-escalatory behaviour in the form of holding people apart (right). Reprinted from Third parties, Violence, and Conflict Resolution: The Role of Group Size and Collective.

This coding resulted in a long behavioural chain, providing information on who did what to whom and in which sequence. Finally, the researchers coded the size of the groups, whether the bystanders present were active or passive, and the outcome severity of the conflict. In total, there were 312 participants across all videos, allowing for statistical analyses of the patterns of behaviours that occurred across the 42 incidents.

Contrary to the idea that groups are detrimental to the social fabric of society, analysis of the CCTV data showed that there was a greater tendency for groups to contribute to de-escalation than escalation. Increased group size was not associated with a difference in levels of antisocial behaviour, but increased levels of de-escalation and prosocial acts. This greater likelihood of third-parties to carry out de-escalatory (rather than escalatory) behaviours

predicted a reduction in violence severity. In contrast to the bystander effect, an increased number of third-parties resulted in a greater likelihood of third-party intervention. A further sequential analysis of the behaviours of those involved in the incidents showed that the chronological pattern of the actions of third-parties—rather than perpetrators or victims—was most significant in predicting whether a conflict escalated to severe violence. More specifically, examination of the triple-interact sequence (Taylor, Donald, & Conchie, 2012)—the first three responses made in response to an aggressive cue—identified that the likelihood of further escalation was best predicted by the third act in the sequence when delivered by a third-party. If this third act was escalatory in nature, then the incident was approximately twice as likely to end in severe violence. However, if this act was de-escalatory, then the most likely outcome was no or minimal violence.

Together, these results show the importance of social context and cumulative peer response in reducing violence, and suggest that within the domain of NTE violence, successful intervention comes from the coordinated action of the group. The central finding is that collective self-regulation is taking place. Third-parties are heavily involved in creating the conditions necessary for violence reduction, and they employ pacifying behaviours aimed at achieving this state. Even in the absence of any formal figures of authority, groups and third-parties have shown to be effective ‘policers’ who can shape the trajectory of violence and encourage peace. These results dovetail nicely with the sociological perspective of informal social control and provide systematic empirical support of how explicit social control is carried out at a micro-interactional level in vivo. The importance of third-party action in the reduction of conflict escalation also runs counter to Felson’s (1993) assumption that third-parties largely reinforce, or are impassive to, the aggressive behaviours of others.

2.2.1 CCTV Limitations

Although CCTV data are a valuable resource for capturing real-life aggressive interactions, such data also have several limitations that must be considered. One crucial

limitation is that most CCTV footage does not contain sound, which restricts coding to only observable acts. Additionally, CCTV footage provides limited information on the identity of the individuals and their personal characteristics. It is also impossible to ascertain the previous history between the individuals in the clips and to gauge a full understanding of the events that have led to the conflict. Similar to naturalistic observations, researchers using CCTV footage are still restricted in what they can see. At times, the actions of those recorded in the videos are unidentifiable, owing to low resolution, awkward filming angles or poor camera focus. There is also the chance of people or obstacles blocking the cameras' view.

Aside from these technical limitations, researchers also need to consider whether the footage covers the event in its entirety, or if the recording commenced too late or ceased too early (Nassauer & Legewie, 2016). Failure to capture the whole event means that valuable interactions between participants are lost. This restricts analyses to particular fragments of the unfolding dynamics. Further, as the CCTV footage tends to be collected en masse and then collated and archived by a team external to the researcher (e.g., a city council, municipality or police force), there is also the possibility that the footage may be overedited, incomplete or exclude behaviours that a third-party deemed not relevant. These exclusion criteria can mean that a large CCTV corpus may be reduced to a much smaller dataset; for example, Levine et al.'s (2011) corpus of 200 videos was reduced to 42 working videos for analysis.

There is also the question of how reactive participants are to CCTV surveillance, and whether those being recorded are altering their behaviour in response to its presence. This remains an open debate in the field. While laboratory evidence suggests that people alter their altruistic behaviours while under the surveillance of cameras (van Bommel, van Prooijen, Elffers, & van Lange, 2014), it can also be argued that the omnipresence of cameras, and the 'normality' of being recorded in public spaces, means that behavioural reactivity to CCTV observation is low. This habituation thesis is supported by research that has found no evidence

that CCTV presence deters violent crime in public spaces (Welsh & Farrington, 2002, 2009). However, additional research is required to settle this dispute.

For researchers interested in using this type of ‘real-life’ data, there are also several ethical deliberations that need to be taken into consideration. Major ethical issues relate to anonymity, confidentiality and consent. Owing to this, it is important that researchers wanting to use CCTV footage collected from public spaces for data analysis have carefully considered and respected the privacy of the individuals being studied. As those observed are unable to provide consent, the camera data should only be sourced from locations where those captured might expect to be observed by strangers. These areas must be public spaces, and, for clarity, must be clearly demarcated from other areas that may be confused for ‘private’ spaces. Further, the presence of a CCTV system must be clearly signed, and the cameras should not be hidden or obscured. To assure confidentiality and anonymity, it is imperative that all CCTV footage is safely stored in a secure location. There should be no existing information that identifies the individuals in the footage, and no such information should be sought by the researchers. As the video footage (and any images pertaining to the footage) is being used for research purposes, these images should not be made available to unrelated third-parties. Finally, any still frames being used in academic work should have all identifiable faces pixelated to protect their identities.

While these suggestions assist in promoting good ethical practices, researchers may still face difficulty in obtaining CCTV data for their work. Importantly, data protection and privacy rules vary from country to country and from organisation to organisation, which means that researchers must be aware of the laws governing their local setting. Further, one must consider the individual concerns of those responsible for the recording and handling of the CCTV footage. First, the management of such footage is labour-intensive and financially costly. Second, it can be an uncomfortable experience for those responsible for the footage to hand it over to researchers, especially if there has been no previous contact or connection

between the parties. Researchers interested in acquiring footage are therefore encouraged to consider the concerns of those agencies, and to spend time building trusting relationships and allaying any concerns with practical solutions. This cannot be achieved instantly, and requires time and effort (this took several years in the case of Levine et al.'s, 2011, association with the city council of North-West England). Providing previous examples of good ethical practice and applied research outputs can also assist with this endeavour.

2.3 Study 1: A CCTV Microanalysis of Within-Group and Between-Group Acts During Aggressive Encounters

The work of Levine et al. (2011) was novel in that it analysed real behaviours occurring during public space violence. The findings, which stress the importance of third-parties as explanatory components of violence, also champion a reappraisal of dyadic models of aggression that exclude such considerations. However, one limitation of the behavioural microanalysis of Levine et al. (2011) is that the researchers did not distinguish the social group memberships of those involved. Rather, the term 'group' was applied to denote any number of bystanders exceeding one. Therefore, the study provided no information on whether an aggressive episode occurred between people who were of the same group and knew each other (e.g., friendship groups), or unknown stranger groups. Further, the study provided no information on whether each action that occurred was directed towards someone of the same social group, or a person of a different social group. This quality of relationship (intragroup interactions as opposed to intergroup interactions) could have a bearing on the potential rate of intervention and the likelihood of peace-making behaviours.

For example, in the work of Levine et al. (2011), it may be the case that third-parties intervened and tended to de-escalate a fracas because those involved belonged to the same social group and wished to maintain group relations, preserve group stability and enhance future collective cooperation (de Waal, 2000). Conversely, for conflicts between two opposing groups in public spaces, there would be a limited (or no) expectation of further interactions or

future collective cooperation, and thus less motivation to intervene. Certainly, years of research on group dynamics suggests that social group categorisation may be important in shaping perceptions and subsequent behavioural tendencies (Otten, 2009). Specifically, previous empirical evidence suggests that individuals hold a favourable bias towards ingroup members over outgroup members (Brewer, 2001; Hewstone et al., 2002; Levine, Prosser, Evans, & Reicher, 2005; Mullen, Brown, & Smith, 1992), and offer more help and empathy towards one's own over 'others' (Lindgaard et al., 2017; Slater et al., 2013; Stürmer, Snyder, Kropp, & Siem, 2006; Tarrant, Dazeley, & Cottom, 2009). This empathy-helping ingroup bias becomes especially relevant in times of direct competition (Richins et al., under review). Therefore, unless ingroup deviants or intragroup conflict threaten the values and content of a particular social group (Marques et al., 2001; Stott et al., 2001; Vidmar, 2001), one would expect a greater likelihood of pacifying intervention and a higher rate of de-escalatory peace-making behaviours towards ingroup members in conflict than outgroup members—in short, we would expect an intragroup de-escalatory bias.

Not only may one expect a de-escalatory bias towards one's own group members, but one may also expect a hostility bias towards outgroup members over ingroup members. Past research suggests that individuals are more likely to both distrust and display prejudice against outgroup members over ingroup members (Insko, Schopler, Hoyle, Dardis, & Graetz, 1990), particularly when in conflict. This may lay the foundations for outgroup dehumanisation and further intergroup conflict escalation (Brewer, 2001; Leyens et al., 2000). Further, while social categorisations alone are not directly related to aggression per se, perceptions of ingroup–outgroup conflict are strongly related to aggression towards the opposing 'other' (see realistic conflict theory D. T. Campbell, 1965; Struch & Schwartz, 1989). Beyond favouring one's own over others, an attack on a group member from an 'outsider' can trigger a joint defensive response and a greater willingness of violent retaliation (Decker, 1996; Tedeschi & Felson, 1994; Vasquez et al., 2015).

Defensive responses to an outgroup are not restricted to situations of intergroup physical threat; rather perceived identity threat from an outgroup—such as having one’s own group unfairly and indiscriminately misrecognised, negatively characterised, or unfairly stereotyped—can shift ingroup norms and identity content to favour outgroup hostility (Reicher, 1996; Stott, Adang, Livingstone, & Schreiber, 2008; Stott, Scothern, & Gorringer, 2013). The elaborated social identity model (ESIM; Drury & Reicher, 2000; Reicher, 1996) provides an interactive intergroup framework of this phenomenon and sets the conditions to describe its process. In the ESIM model, psychological subgroups can exist within one physical aggregate group (for example, subgroups of environmental protestors campaigning at a rally, or hooligan football fans present with non-hooligan fans of the same team during a sporting event; see the protest work of Drury & Reicher, 2000, or the World Cup Italia 90 football fan discussion of Stott et al., 2008). These different subgroups may be unfairly perceived by an external group (such as the police) as a single, indistinguishable entity that endangers social order. For the majority of ‘moderates’, this unfair and indiscriminate treatment limits the freedom to conduct socially legitimate activities (such as the ability to engage in a peaceful protest or to sing loudly as to support one’s team).

Driven by a shared sense of victimhood, and often as a consequence of experiencing coercive actions from the more powerful group who attempt to impose their own definition of legitimacy (such as through indiscriminate use of tear gas, or baton charges), these ‘moderate’ subgroup members may recategorise with the more extreme subgroups under a single collective identity. Recategorisation may result in a shift of category prototypicality and group norms to radical positions that support retaliation and intergroup hostility as appropriate actions. Here, group violence escalation emerges from the dynamic interplay of perceived (il)legitimacy of the wider intergroup context. In this sense, group identities and normative behaviour are not fixed in advance, but instead alter as a function of intragroup and intergroup interactions and perceptions (Drury & Reicher, 2000; Postmes, Haslam, & Swaab, 2005).

Despite empirically-derived intimations of higher levels of intergroup hostility and intragroup pacification during aggressive events, research into these potential biases during naturalistic violent conflicts is scant. One exception is the recent work of Liebst, Heinskou and Ejbye-Ernst (2017), who analysed CCTV footage of violent street fights in Copenhagen to ascertain, via logistic regression analyses, the predictive factors for intervening third-parties being victimised by an aggressor. The authors coded the observable interventions (whether it was an escalatory or de-escalatory bystander intervention) of 229 intervening bystanders across 64 clips. Subsequently, the authors recorded whether these bystanders were met with physical victimisation. In addition, measurements were recorded regarding the severity of the conflict, number of bystanders present, gender of the bystander and group membership of the bystander. Shared group membership was classified according to whether the bystander was of the same friendship or acquaintance group as the intended initial target. This was determined through the coders' observation of certain social relationship cues between the bystander and victim for each video, including a shared focus of attention and interpersonal distance (Ge, Collins, & Ruback, 2012; Liebst et al., 2017; Murphy, 2016). The social group codes for each bystander in the clips were then validated against the corresponding police case file that provided written descriptions of each incident. The authors predicted that perceptions of intergroup conflict would heighten intergroup derogation and target legitimacy. Therefore, if an intervening bystander shared social group membership with the initial target (and thus had outgroup status from the aggressor), then that bystander would have an increased likelihood of victimisation upon intervention.

Results from the logistic regression estimation of the data found that an intervening bystander with shared group membership to the target was over seven times more likely to be violently victimised than bystanders of a differing social group (which included interveners from the aggressor's social group). This 'social' variable was the best predictor of victimisation, followed by the intervention strategy used by the bystander (escalatory rather

than de-escalatory). The variables of gender, severity of violence and number of bystanders were non-predictive of the outcome. This group-based victimisation bias was attributed to social categorisation processes. More specifically, the authors proposed that intergroup hostilities between a perpetrator and the target led to categorisations of that opponent as an outgroup ‘other’. Those closely related to the target were then also categorised by the perpetrator as outgroup ‘others’ and, as such, attributed with the same negative emotions and target legitimacy (Liebst et al., 2017). As a result, the intervener could also become a target for physical aggression and attacked, an effect that was rarely observed when those intervening were not ‘socially related’ to the initial target.

This central finding stresses the appraisal of social relationships as important explanatory components of public space violence. It also suggests the presence of an intergroup hostility bias in non-experimental data; however, the magnitude of such an effect should be met with caution. One major caveat of this study originates from the data itself. Liebst et al.’s (2017) CCTV sample consisted entirely of footage provided by the Copenhagen Police Department. A major advantage of using police footage is that each clip is accompanied with a corresponding criminal file that can be used to confirm the social group identity and familiarity between participants—something that is not measured in Levine et al.’s (2011) earlier work. However, the disadvantage of using police data is that the data have been selected on the dependent variable: cases that have been reported to the police that are usually severe. Therefore, the sample may greatly over-represent severe outcomes—particularly intergroup conflicts that may have been more likely to be reported to the police, as opposed to intragroup conflicts that may have been settled internally.

Unlike Liebst et al.’s (2017) work that relied on footage from static, unstaffed cameras, Levine et al.’s (2011) original CCTV data was acquired from UK city council camera operatives. These operatives were instructed to proactively record any interactions that looked like they might escalate into a conflict, up to the point where there was either

natural conflict dissipation or police intervention. The sample therefore contained conflicts of varying severity, from disputes that involved merely posturing and blustering through to conflicts that resulted in severe violence. This meant that Levine et al. (2011), arguably, had a more representative sample of public space violence, with incidents varying in severity. Further, as Levine et al.'s (2011) data were not sampled on whether they were reported to the police, they also presumably contained a more accurate apportionment of intragroup conflicts. As a result, we may expect the large effect of intergroup hostility ($OR = 7.7$) reported in Liebst et al.'s (2017) study to be more inflated than the effect in an alternative CCTV sample (as obtained by Levine et al., 2011) that does not have the same police-reported sample selection bias. It is important to note that the CCTV data we have analysed is a subset of Levine et al.'s (2011) original collected data.

In spite of these sampling considerations, it must be emphasised that an accompanying police file is undoubtedly a qualitatively-rich asset. Not only did it provide Liebst et al. (2017) with information on the history and circumstances of each incident, but it also provided a reliable reference to ratify whether social groups in public space violence could be reliably coded 'a priori' from behavioural cues alone. On this point, Liebst et al. (2017) stated:

Here, however, we should add that our video-based assessment of relationships rarely¹ had to be corrected when compared to the police case file descriptions. This stresses the validity of our social group measure and is, furthermore, encouraging for scholars who have an interest in bystanders' social relationships but no access to other data sources to validate their video-based assessment of these relations. In our experience—and consistent with evidence suggesting that social group characteristics can be accurately inferred

¹ 'Rarely' amounts to a discrepancy between blinded social group codes and the police case file social group information in one of the 69 videos.

from nonverbal cues (Ge et al., 2012; Murphy, 2016)—relationship ties can be reliably coded from video recordings of natural conflicts. (pp. 18–19)

Therefore, while one might expect a limitation of assigning group memberships based on merely viewing the entire CCTV clip, which may unwittingly provide coders with heuristics based on the outcome variable (e.g., friendly or hostile interactions), Liebst et al.'s (2017) work *does* suggest that blind coder interpretation of social groups in videos of public space violence is a reliable process. The high success rate of matching the coders' interpretations of social groups to the information provided in the police files may, therefore, negate the immediate need for police files for social group recognition. This presents an opportunity for researchers to code for social group membership in their videos, and to potentially consider this an explanatory variable for a particular research question, even if self-reported group information is not available. The CCTV data we have obtained for this study falls into this latter category.

Therefore, the aim of this study is threefold. Similar to Levine et al. (2011), our first CCTV study examines the occurrence of escalatory and de-escalatory behaviours by those present in aggressive public space encounters. However, unlike the original work, the current study also aims to code for social groups, thus complying with Levine et al.'s (2011) original recommendation to analyse events at a social group level. By coding social groups, we intend to ascertain whether a de-escalatory bias towards one's own group exists, as predicted in the literature.

The second objective is to extend the victimisation work of Liebst et al. (2017) and confirm whether an escalatory bias exists towards outgroup members in public space violence when using an alternative CCTV dataset not collated for police evidence (i.e., with a city council CCTV sample that is less biased towards severe conflicts and intergroup conflicts). Any notable effects of the 'group' on escalation and de-escalation will provide further support

for the need to consider social group membership as a unit of analysis in models of aggression and violence in public spaces.

Our final goal is to determine whether there is a relationship between the number of active bystanders in a conflict and the level of escalatory and de-escalatory behaviours. This investigation may provide evidence into whether violence research in NTE settings should place emphasis on individuals outside the aggressive dyad. However, unlike Levine et al. (2011), we intend to investigate if the relationship is dependent on conflict type—intergroup versus intragroup conflict.

2.3.1 Hypotheses

Based on the assumption of de Waal (2000) that intragroup conflicts will likely be stifled to preserve group relations, and given research in the social identity tradition that suggests that ingroup members offer more help and empathy towards their own over ‘others’ (Reicher & Haslam, 2009; Slater et al., 2013; Stürmer et al., 2006; Tarrant et al., 2009), we predict:

H1: There will be a proportionally higher rate of de-escalatory acts occurring within groups in conflict than between groups in conflict (intragroup de-escalatory bias).

Given that social categorisations of ‘us’ versus ‘them’ during conflict is strongly related to derogation towards the opposing ‘others’ (Decker, 1996; Liebst et al., 2017; Struch & Schwartz, 1989; Tedeschi & Felson, 1994; Vasquez et al., 2015) we predict:

H2: There will be a proportionally higher rate of escalatory acts occurring between groups in conflict than within groups in conflict (intergroup hostility bias).

In an extension of the intragroup de-escalatory bias and intergroup hostility bias, we predict:

H3: There will be a positive relationship between the number of active bystanders and number of de-escalatory acts for intragroup conflicts (H3a), and a positive relationship between the number of active bystanders and number of escalatory acts for intergroup conflicts (H3b).

2.3.2 Method

Data and sample. The dataset comprises a corpus of CCTV clips capturing aggressive conflicts recorded in British public spaces. This corpus was collated by Levine et al. prior to their 2011 article and contained 384 raw clips (200 clips from the 2011 article, and an additional 184 clips that remained on compact discs and that had not previously been archived). All clips were obtained by permission from two city councils in North-West England and were recorded by qualified camera operatives under employment of those city councils. Each conflict was recorded from the point at which the operative noticed aggressive gesturing or agitation between members of the public to the point when the incident had either naturally dissipated or the police had intervened.

We first excluded the 200 clips from Levine et al. (2011), which left a dataset of 184 clips of varying quality and content. Excess clips of unsatisfactory quality and insufficient content were then excluded leaving a final sample of 34 clips. A clip was determined to have unsatisfactory quality if the camera focus, resolution or frame rate were too low to allow for the systemic coding of participant behaviours. A clip was defined as having insufficient content if the clip broke off during or before the conflict had concluded, if those present exited the camera view or if the clip did not contain a confrontation (e.g., clips that showed offenders being followed on camera to assist with post-conflict police arrests). A clip was also deemed to have insufficient content if it was a duplicate (e.g., two camera operatives had burned the same incident on separate compact discs) or if the clip contained too few instances (< 20) to sufficiently analyse.

To supplement the 34 clips, nine randomly selected clips were also included from the Levine et al. (2011) paper. This brought the number of clips to be coded and analysed to a total of 43. Although we did not originally intend to use clips from Levine et al.'s (2011) study, we included the additional nine clips because it provided an opportunity to validate whether our coding was comparable to previous work that had been done. Specifically, it allowed for inter-rater reliability tests to be conducted on a sample of 20 % ($N = 9$) of the total coded ($N = 43$) clips.² Further, these inter-rater reliability checks could be conducted against the work of an experienced coder with a high level of coding expertise, who had undertaken Observer XT software training and was experienced in the microanalysis of CCTV footage, and who had utilised the same coding scheme as the current study. As our chief variable of interest was group membership, which Levine et al. (2011) did not code, we perceived the 20 % overlap to be less of a concern.

The final sample was dated between January 2005 and October 2008, and the duration of the clips lasted between one and eight minutes. The 43 clips contained a total of 332 participants. All of the 43 incidents involved a protagonist who drove the initial aggression, and a target who received that initial aggression (total $N = 86$). All of the 43 incidents also involved at least one 'active' bystander who had performed at least one escalatory or de-escalatory behaviour towards another person in the clip, or who had received at least one such action. The total number of active bystanders across the 43 incidents was 246.

Of the 43 incidents, 3 incidents contained one active bystander (7.0 %), 5 contained two active bystanders (11.6 %), 3 contained three active bystanders (7.0 %), 6 contained four active bystanders (14.0 %), 3 contained five active bystanders (7.0 %), 5 contained six active bystanders (11.6 %), 8 contained seven active bystanders (18.6 %), 3 contained eight active

² Additional N determined as 43 videos: 20 % = 34.4 videos; 43 videos – 34 videos = 9 additional videos. No information related to coded actions, participant roles, participant genders or severity of incident was provided to the thesis author until after all of the 43 videos had been coded by the author and inter-rater reliability tests could take place.

bystanders (7.0 %), 2 contained nine active bystanders (4.7 %), 1 contained ten active bystanders (2.3 %), 1 contained eleven active bystanders (2.3 %) and 3 contained twelve active bystanders (7.0 %). The mean number of active bystanders per incident was 5.72 ($SD = 3.06$).

The gender composition showed that 34 (79.1 %) of the 43 incidents were initiated by two males, while 9 (20.9 %) incidents were initiated by two females. From the 246 active bystanders, 171 (69.5 %) were males and 75 (30.5 %) were female.³

Coding procedure. Coding of the visual data consisted of three phases: identification of actor roles, recording of individuals' behaviours and towards whom, and assignment of social groups. In the initial coding phase, the thesis author identified the protagonist, target and active bystanders present in each of the 34 initial CCTV clips. As with Levine et al. (2011), the protagonist was defined as the person driving the initial aggression towards a target (see Appendix). In contrast, the target was defined as the person to whom the main protagonist's aggressive behaviours were originally orientated. Active bystanders were classified as all other people who were 'actively' involved in the episode (i.e., they performed at least one behaviour of coding interest towards anyone else in the clip, or who had received at least one such action). Additionally, each bystander was assigned a unique numeric identity (ID) code to ensure that any of their actions or receipt of an action could be recorded in future coding.

³ Across the 43 videos, the 93 active female participants contributed a total of 976 actions. Of these actions, 38.9 % (380) were escalatory and 61.1 % (596) were de-escalatory. Across the 43 videos, the 239 active male participants contributed a total of 2579 actions. Of these actions, 37.1 % (956) were escalatory and 62.9 % (1623) were de-escalatory. A binomial logistic regression with a random intercept term to control for overdispersion (see Warton & Hui, 2011) found no statistical effect of gender on the weighted proportions of escalatory and de-escalatory behaviours recorded across the 332 active participants ($\chi^2(3, N = 332) = 0.74, p = .391, n.s.$). Therefore, when interacting with others, females' actions were just as likely to be escalatory or de-escalatory as males. Although these results are extremely interesting within themselves, as gender was not a central focus of this thesis and no effect of gender was found, in the interest of brevity, we no longer consider gender as an explanatory variable in the remainder of this thesis.

In the second coding phase, the thesis author separately coded the observable actions of each of these actors across the initial 34 clips and recorded towards whom each behaviour was directed. Similar to Levine et al. (2011), each discrete action was assigned to one of three behavioural categories: a physical escalatory act, a non-physical escalatory act or a de-escalatory act (see Appendix). Physical escalatory acts included pushes, slaps, punches, kicks, scratches, chokes, grappling and headlocks. Non-physical escalatory acts included aggressive hand and body gestures, aggressive pointing, feinting, invasion of space and the removal of one's own clothes. In the same manner as Levine et al. (2011), these two escalatory categories (physical and non-physical) were subsequently aggregated into a single category: 'escalatory acts'. De-escalating behaviours included calming open-hand gestures, standing between people who were confrontational, pulling an aggressive person away, walking away from the fight and covering one's self or others from blows. This resulted in a behavioural chain for each video, which specified at what time in the clip a person did an action to another.

In addition, supplementary variables, including the gender of each actor and the severity of violence in the clip (mild, moderate or severe), were coded. Violence was considered mild if there was some posturing and gesturing, light shoving and feinted movements, but no physical punches, kicks or slaps that were landed on an opponent. Violence was considered moderate if a physical fight took place, but nobody was badly hurt, knocked out or kicked repeatedly. Violence was considered severe if a person was badly hurt, knocked unconscious, kicked severely while on the floor or bleeding profusely.

The thesis author then repeated these coding steps with the additional nine video clips that were included in the Levine et al. (2011) study.

To determine the reliability of coding, we compared the thesis author's codes for the subsample of the 9 (21 %) clips to those of Levine et al. (2011). For the first phase of coding (the assignment of actor roles), agreement was based on whether the two coders had selected the same individual as the protagonist and the same individual as the target. The two coders

had a high level of agreement on the assignment of the specific actor roles within these confrontations (Krippendorff's $\alpha = .88$). For the second phase of coding (observable actions), agreement was defined as a match between coders when specifying the actor, the behavioural action, and the recipient. Agreement between the coders for observable actions was sufficient (Krippendorff's $\alpha = .80$). Agreement on the additional variables of gender and outcome severity measures were high (Krippendorff's $\alpha = .87$).

In the final phase of coding, two previously unused coders, blinded to the purpose of the study, separately determined whether the conflicts witnessed in the CCTV clips were intragroup, intergroup or unknown. Similar to Liebst et al. (2017), shared group membership was ascertained by viewing the clip in its entirety, including any periods captured before and after the conflict, and noting if the fighters appeared to be cohesive and familiar with one another. This was determined by observing behaviours such as travelling around the streets together, linking arms, wearing similar fashion or fancy dress, and appearing to share friends and acquaintances determined via a shared focus of attention and interpersonal distance (Ge et al., 2012; Liebst et al., 2017; Murphy, 2016). Intergroup conflicts were determined by whether the fighters appeared to be unfamiliar to one another or of different social groups. In total, the coders agreed that 20 of the clips displayed intragroup fights and 23 of the clips displayed intergroup fights (Krippendorff's $\alpha = .86$).

Once clips had been classified according to whether the conflict was intergroup, intragroup or unknown, the two coders together assigned each of the actors (protagonist, target, bystander 1 and bystander 2, etc.) in each clip a group number. For example, if the protagonist, target and bystanders were of the same social group, they were all assigned a shared group number. If they appeared to belong to different groups, they were assigned a separate group number. This allowed us to later determine if actions were aimed towards someone of the same social group (intragroup action) or aimed towards someone of a different social group (intergroup action).

2.3.3 Results

The majority of the 3555 actions coded across the corpus of 43 CCTV clips (63.1 %, 2243 out of 3555; $\chi^2(1, N = 3555) = 243.8, p < .001$, Cramer's $V = 0.26$) were de-escalatory in nature (e.g., calming open-hand gestures, de-escalatory touching and movements to block contact). The remaining 36.9 % (1312) of behaviours were escalatory (e.g., punching, kicking, grappling, pushing, aggressive gesturing and pointing, and invading of space).

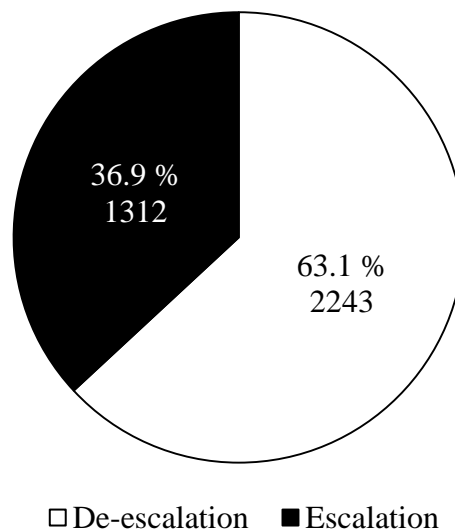


Figure 2.2. Total number of de-escalatory and escalatory acts as a percentage of the sum of all behaviours ($N = 3555$)

Of the 3555 actions recorded, the majority (66.3 %, 2357; $\chi^2(1, N = 3555) = 1760.65, p < .001$, Cramer's $V = 0.50$) were performed by bystanders. 20.1 % (713) of actions were performed by protagonists and 13.6 % (485) of actions were performed by targets (with a statistical difference between these roles, $\chi^2(1, N = 1198) = 43.39, p < .001$, Cramer's $V = 0.19$).

The majority of the 2357 bystander actions recorded (77.6 %, 1829; $\chi^2(1, N = 2357) = 718.12, p < .001$, Cramer's $V = 0.55$) were de-escalatory. 22.4 % (528) of bystander actions were escalatory. Third-party bystanders therefore, have a tendency to attempt to de-escalate conflict.

The majority of the 713 protagonist actions recorded (83.3 %, 594; $\chi^2(1, N = 713) = 316.45, p < .001$, Cramer's $V = 0.67$) were escalatory. 16.7 % (119) of protagonist actions were de-escalatory. Protagonists therefore, largely attempt to escalate conflicts.

The majority of the 485 target actions recorded (55.9 %, 271; $\chi^2(1, N = 485) = 6.70, p = .010$, Cramer's $V = 0.12$) were de-escalatory. 44.1 % (214) of bystander actions were escalatory. Targets, by and large, try to de-escalate conflict, but almost half of their actions are escalatory.

Of the total 3555 behaviours recorded, the majority (58.3 %, 2073 out of 3555; $\chi^2(1, N = 3555) = 98.25, p < .001$, Cramer's $V = 0.17$) were directed towards an ingroup member, with 41.7 % (1482 out of 3555) aimed towards an outgroup member.

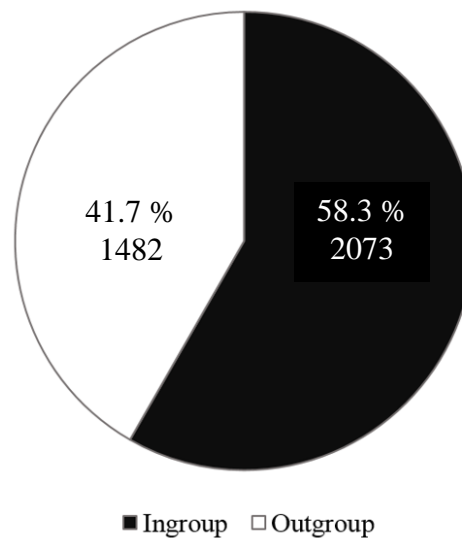


Figure 2.3. Total number of intragroup and intergroup acts as a percentage of the sum of all behaviours ($N = 3555$)

At face value, the higher number of actions directed towards ingroup members over outgroup members suggests an intragroup interaction bias. However, it is plausible that participants on average across all videos had greater exposure to ingroup members than outgroup members, and thus potentially a higher likelihood of intragroup interaction.

Specifically, a video capturing an intragroup dispute typically involved mainly (if not entirely) one group. This meant that the majority (if not all) of the actions in an intragroup video could be exclusively intragroup. In comparison, an intergroup dispute captured both intergroup actions between quarrelling groups, and also intragroup actions as those within their own groups internally negotiated the situation. This increased ingroup exposure is descriptively suggested in the comparison of the total count of videos that contained at least one intragroup action (all 43 videos) versus the total count of videos that contained at least one intergroup action (31 of the 43 videos).

To account for this potential ‘exposure effect’, and also for variations in the length of videos (longer videos typically contained more actions; Pearson’s $r(43) = .85, p < .001$), the total number of intergroup-directed actions (de-escalation and escalation) and intergroup-directed actions (de-escalation and escalation) for each video were converted into proportions. Intragroup behavioural proportions were attained by dividing the number of intragroup de-escalatory acts and intragroup escalatory acts in each of the individual 43 videos by the total number of intragroup acts in that video. In a similar fashion, intergroup behavioural proportions were attained by individually dividing the number of intergroup de-escalatory acts and intergroup escalatory acts in each of the 31 videos (containing intergroup behaviours) by the total number of intergroup acts in that video. This provided a total of 74 cases⁴ for data analysis: 43 video cases depicting intragroup escalatory and de-escalatory behavioural proportions, and 31 video cases depicting intergroup escalatory and de-escalatory behavioural proportions.

To identify general trends in the proportional compositions of de-escalation and escalation across videos and group conditions, these data were visualised in two box plots. In

⁴ It is important to note that intragroup actions and intergroup actions could be coded within the same video. This may present issues regarding the non-independence of data; a point expanded upon in the discussion of this study.

these plots (see Figures 2.4 and 2.5) each point represents a video; black points depict the intragroup-directed actions that occurred within a video ($N = 43$), while white points depict the intergroup-directed actions that occurred within a video ($N = 31$). Initial inspection of the scatter of points and corresponding box plot whiskers suggested large variation in the proportion of actions between cases; however, this scatter was larger for intergroup cases than intragroup cases. In Figure 2.4, intragroup cases generally appeared higher up the de-escalation axis ($M = 0.75$, $SD = 0.21$) than intergroup cases ($M = 0.55$, $SD = 0.26$). Correspondingly, in Figure 2.5, intergroup cases generally appeared higher up the escalation axis ($M = 0.45$, $SD = 0.26$) than intragroup cases ($M = 0.25$, $SD = 0.21$). Put simply, there appeared to be a proportionally higher level of de-escalatory actions within groups than between groups, and a proportionally higher level of escalatory actions between groups than within groups. De-escalation was present in each of the 74 cases.

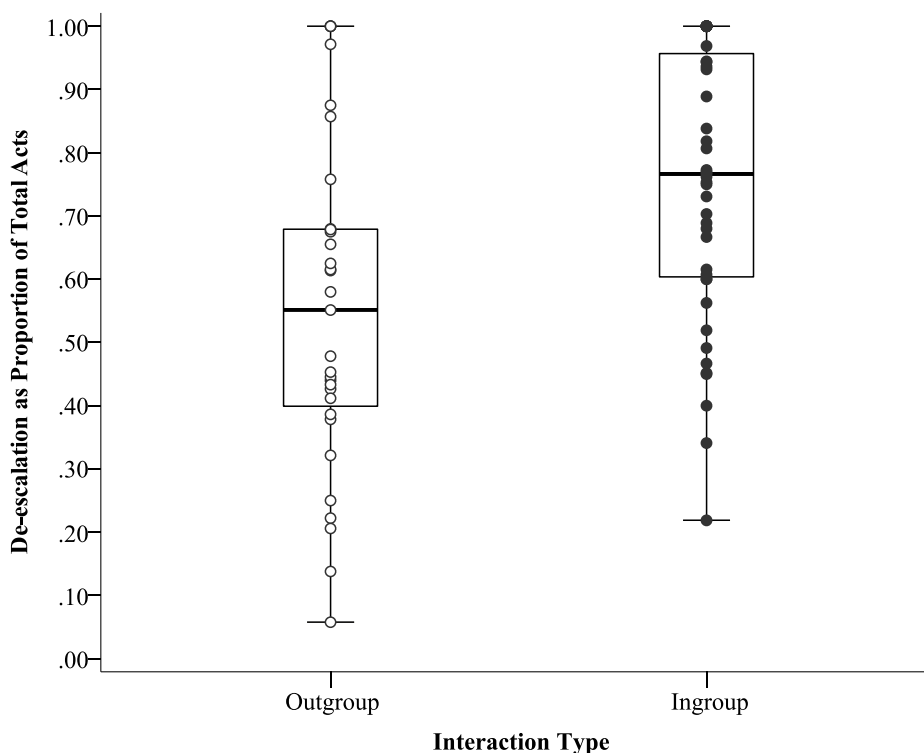


Figure 2.4. De-escalation as a proportion of intergroup total acts and intragroup total acts for each video ($N = 74$)

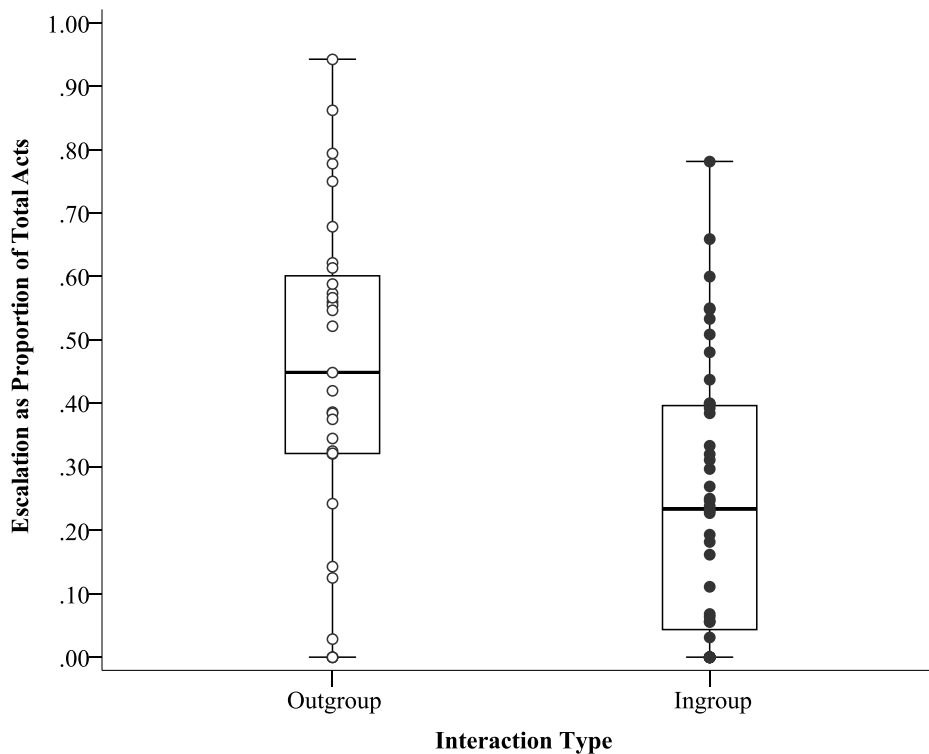


Figure 2.5. Escalation as a proportion of intergroup total acts and intragroup total acts for each video ($N = 74$)

A binomial logistic regression of proportions (Warton & Hui, 2011) was conducted to ascertain whether the average weighted⁵ proportions of de-escalation and escalation differed statistically at the group level when controlling for several covariates. The first covariate to be included was the number of active individuals present in a clip. This covariate was deemed important because there are empirical suggestions that increased bystander presence is related to increased violence (Felson, 1982; Jankowski, 1991; W. B. Sanders, 1994; Tomsen, 1997; Weenink, 2014) or decreased interventions (Latané & Darley, 1970), and, conversely, increased levels of conciliatory actions and no increase in escalatory acts (Levine et al., 2011). The second covariate included was the severity of violence. This was included because there

⁵ To retain information for a sample size, all proportions were weighted by the total number of acts from which the proportion was estimated. This ensured greater confidence in a proportion derived from a larger sample size (e.g., a proportion of 75 % attained from 75 de-escalatory acts out of 100 acts in a video) than those proportions derived from a smaller sample size (e.g., a proportion of 75 % attained from three de-escalatory acts out of four).

was a possibility that a more severe incident could include higher levels of escalation and potentially decreased levels of de-escalation (as people did not act to prevent the violence) or increased levels of de-escalation (as people intervened more to stop the severe violence; Parks et al., 2013). Finally, since there may be a higher opportunity for intragroup interactions in videos depicting intragroup conflicts, and a higher opportunity for intergroup interactions in videos depicting intergroup conflicts, the type of conflict (whether primarily an intragroup conflict or intergroup conflict) was included as a final covariate. In addition to these covariates, the logistic regression model contained a random intercept term to account for overdispersion (see Warton & Hui, 2011).

The binomial logistic regression comparing the de-escalatory and escalatory proportions of intragroup actions and intergroup actions when controlling for covariates was statistically significant ($\chi^2(6, N = 74) = 20.53$, 95% CI [0.95, 1.96], $p < .001$, odds ratio = 4.23). Intragroup interaction cases contained a significantly higher proportion of de-escalatory acts than did intergroup interaction cases. Accordingly, intergroup interaction cases contained a significantly higher proportion of escalatory acts than did intragroup interaction cases ($\chi^2(6, N = 74) = 20.53$, 95 % CI [0.95, 1.96], $p < .001$, odds ratio = 0.24).

An examination of the behavioural chain (see Table 2.1) showed that after an act of intragroup escalation, the most likely response (52.3 %) was intragroup de-escalation ($z = 3.43$, $SE = 0.04$, $p < .001$). In 37.4 % of occasions, intragroup escalation was succeeded by another act of intragroup escalation. Comparatively, after an act of intergroup escalation, the most likely response (46.7 %) was further intergroup escalation ($z = 3.60$, $SE = 0.04$, $p < .001$). In 31.3 % of occasions, intergroup escalation was succeeded by intergroup de-escalation, and in 19.7 % of occasions, intergroup escalation was succeeded by intragroup de-escalation.

Table 2.1

The likelihood of the next sequential action succeeding an act of intragroup escalation, intragroup de-escalation, intergroup escalation and intergroup de-escalation, N = 3512

Action	Next Intragroup Action		Next Intergroup Action		Sum (N)	
	Escalation	De-escalation	Escalation	De-escalation		
Intra- group	Escalation	37.4 % (226)	52.3 % (316)	2.2 % (13)	8.1 % (49)	100 % (604)
	De-escalation	21.1 % (303)	57.7 % (830)	9.8 % (141)	11.5 % (165)	100 % (1439)
Inter- group	Escalation	2.3 % (16)	19.7 % (138)	46.7 % (328)	31.3 % (220)	100 % (702)
	De-escalation	6.5 % (50)	21.6 % (166)	27.5 % (211)	44.3 % (340)	100 % (767)

Relationship between number of bystanders and behaviours. Partial correlations were performed to determine the relationship between the number of active bystanders and number of de-escalatory and escalatory acts in a video while controlling video length. For videos depicting intergroup fights, there was no relationship between the number of bystanders and number of escalatory acts ($r(20) = .06, p = .777, N = 23$), and a significant positive partial correlation between the number of bystanders and number of de-escalatory acts, when controlling for video length ($r(20) = .69, p < .001, N = 23$).

For videos depicting intragroup fights, there was a significant negative partial correlation between the number of bystanders and number of escalatory acts, and a significant positive partial correlation between the number of bystanders and number of de-escalatory acts, when controlling for video length ($r(17) = -.62, N = 20, p = .004$, and $r(17) = .51, N = 20, p = .026$, respectively).

2.3.4 Discussion

This study examined the occurrence of intragroup-directed and intergroup-directed escalatory and de-escalatory behaviours by those present in public space conflicts. Through the microanalysis of CCTV footage, we aimed to empirically examine the theoretically-derived existence of both a de-escalatory bias towards ‘one’s own’ and an escalatory bias

towards social ‘others’ in examples of public space violence. Unlike the most comparable CCTV analysis study of Liebst et al. (2017) that examined the importance of social group membership on likelihood of bystander victimisation using footage from police cases, we conducted this research with data from city council-operated cameras and focused on the differences in the type of bystander intervention, rather than cases of victimisation of bystanders.

In line with previous research (Liebst et al., 2017; Parks et al., 2013), the majority of actions coded from the 332 active participants were de-escalatory in nature. After subdivision of these actions at the group level (i.e., whether the actions were directed towards an ingroup member or an outgroup member), the majority of behaviours in each category remained de-escalatory, although a significantly larger proportion of intragroup behaviours were de-escalatory in comparison to intergroup behaviours (H1). Correspondingly, a significantly larger proportion of intergroup-directed behaviours were escalatory in comparison to intragroup-directed behaviours (H2).

The finding that de-escalation is the principal behaviour during conflicts—even in intergroup cases—suggests that public space violence is not one mass melee. Rather, regulation is demonstrated among participants, who, overall, attempt to stifle the conflict. This is in line with ethnographic work that stipulates that public space conflicts that may appear unruly are in fact strongly normalised with ‘rules of disorder’ and implicit social values that prescribe when conflict escalation requires intervention (Fox, 1978; Marsh, Rosser, & Harré, 1978). The high level of de-escalatory acts further provides evidence that violence regulation can occur from both within, and between, groups. These findings are novel and challenge the preconception that groups in conflict are embedded with an ‘us’ versus ‘them’ mentality, and are solely antagonistic. However, it is also important to note that intergroup de-escalation was proportionally less likely than intragroup de-escalation.

The finding that a substantially larger proportion of intragroup actions rather than intergroup actions were de-escalatory suggests the existence of a de-escalatory bias towards ingroup members (supporting H1). This is further supported by the examination of the behavioural sequences that found that ingroup escalatory acts tended to be met with ingroup de-escalation. This ingroup de-escalatory bias may be a behavioural expression of the valued relationship hypothesis (Aureli et al., 2002)—specifically, conflict threatens to deteriorate social relationships and relationships are commodities that must be protected (de Waal, 2000). This interpretation might fit well for familiarity groups, such as friends and acquaintances, who share a social identity from previous co-actions and pre-existing social ties (Koudenburg, Postmes, Gordijn, & van Mourik Broekman, 2015; Prentice, Miller, & Lightdale, 1994; Swann et al., 2012). In this instance, these individuals have a past and can expect to have a future; thus, de-escalation of internal conflicts helps to preserve this continuation. However, the valued relationship hypothesis may be less useful when considering social category groups whose members identify together through a perceived similarity with some (and distinctions from others) on a contextually relevant, salient dimension (Turner et al., 1987). For example, in a scenario where an intervener has a shared salient football-team identity with a stranger protagonist, there exists no previous interpersonal relationship, and, most probably, no expectations of future interactions. As such, any intragroup de-escalatory attempts may be driven by separate considerations, such as metaperceptual concerns that one's social category group may be negatively evaluated by others who are observing (O. Klein & Azzi, 2001; Vorauer, Main, Roy, & Hunter, 2000). Here, motivations for intragroup de-escalation during intragroup conflicts diverges from the emphasis on communal familiarity that is dominant in the non-human primate literature.

It is also important to note that intragroup de-escalations can also occur during intergroup conflicts. Examination of the behavioural sequence found that almost one-fifth of intergroup escalatory acts were succeeded by intragroup de-escalatory acts. De-escalation of

fellow group members during intergroup conflicts may be a behavioural representation of ‘self-policing’ in which group members manage their own deviants, whose escalatory actions may negatively affect the image of the wider social group (Stott et al., 2007, 2001). These intragroup de-escalations during intergroup violence may also reflect an overarching superordinate framework that sets the appropriate limits of intergroup violence, and signals to fellow group members when violence is legitimate and how it should be perpetrated. For example, how the overarching framework of a ‘fair fight’ regulates intergroup barroom fights (Copes et al., 2013), how the rules of a stage audience keep violent contests within socially enforced limits (Collins, 2008), and historically how the chivalric code shaped the rites of religious violence in sixteenth-century France (Davis, 1973). Alternatively, de-escalation of one’s own group members during intergroup conflicts may also be an act of concern, to prevent a fellow group member being physically injured (Levine et al., 2012) or potentially arrested.

Attempting to de-escalate ingroup members, whether during an intragroup fight or intergroup fight, still remains a potentially costly intervention. There is the risk of injury from a spillover of violence and the potential for police arrest by merely being involved in a public fracas. Swann et al. (2012) suggested that such willingness to prosocially engage on behalf of the ingroup, even at great cost, could be explained by the visceral feeling of oneness with the group. Regardless of the motivation, the tendency to de-escalate ingroup members is in line with a growing body of bystander research that shows that shared group membership predicts an enhanced likelihood of placatory intervention (Levine & Manning, 2013; Levine et al., 2005; Lindegaard et al., 2017; Slater et al., 2013).

A significantly larger proportion of intergroup-directed actions were escalatory in comparison to intragroup-directed actions (supporting H2). Further examination of the behavioural sequence of actions found that intergroup escalatory acts were predominately succeeded by further intergroup escalatory acts, whereas intragroup escalatory acts were

predominately succeeded by intragroup de-escalatory acts. Together, these provide behavioural evidence of a hostility bias towards outgroup others in times of conflict. These findings are perhaps unsurprising given the current predictions in the literature. Otten (2009) theoretically suggested that the categorisation of ‘us’ and ‘them’ had an impact on cognitions, motivations and subsequent behaviours related to intergroup hostilities. Specifically, by default, the behaviours of outgroup members in conflict were perceived as more hostile than equivalent behaviours from ingroup members during ingroup conflicts (Otten, 2009). Further, outgroup attacks on an ingroup member may be generalised as an attack on the entire group, prompting increased social unity, increased outgroup entitativity and a greater willingness to aggressively retaliate on behalf of the group (Decker, 1996; Tedeschi & Felson, 1994; Vasquez et al., 2015). Finally, qualitative interviews with young offenders and patrons of the NTE all confirm the existence of an implicit and explicit normative expectation that in times of intergroup conflicts, one should ‘stick up for their own’ (Levine et al., 2012; Weenink, 2014), even if that individual is in the wrong (Graham & Wells, 2003). These escalatory partisan actions, while potentially costly, may be explained by the visceral feeling of oneness with the group (Swann et al., 2012).

While this may explain why an intergroup hostility bias was found, it must be repeated that the most common intergroup behaviour recorded was de-escalation. In fact, each intergroup data point contained at least some cross-group de-escalatory attempts. This poses the interesting question of why there was so much cross-group de-escalation (on average, over half of all actions) when there is a generalised normative expectation to fight an opposing ‘other’, and when the risk of cross-group intervention has an even higher risk of victimisation (Liebst et al., 2017). The most obvious explanation might again be the willingness to intervene on behalf of one’s own group member (i.e., ‘I don’t want my fellow group member to get hurt or arrested, therefore I will attempt to de-escalate the outgroup opponent’). Alternatively, individuals in these violent emergencies may not have appraised the situation as

‘us’ versus ‘them’ at all. Instead, by sharing a position on violence with similar-minded bystanders (regardless of initial social group), individuals may have recategorised themselves through metacontrast: we (who are not violent) against them (who are fighting) (Hogg & Reid, 2006; Turner et al., 1987). Here, it would be a case of ‘*we* really need to do something about *those* fighting’. This may explain—qualitatively observed in the video data, but not specifically reported here—the joint attempts from members of opposing social groups to work together in unison to de-escalate a fight. Finally, microsociologist Collins (2008) looked beyond the group level and suggested that violence was difficult for the human species as a whole. He postulated that “humans are hard wired for interactional entrainment and solidarity [...] We have evolved, on the physiological level, in such a way that fighting encounters a deep interactional obstacle” (p. 27). As such, entrainment and solidarity with others (regardless of social group boundaries) works to prevent violence. This, Collins (2008) claims, is the evolutionary price paid for cooperative living and civilisation. Although all of these suggestions are tentative up to this point, what is conclusive is that a partisan ‘us’ versus ‘them’ rhetoric is oversimplified and ignores the complexities of the real-life intergroup placatory behaviours observed.

A final finding of potential interest is the manner in which escalation tended to be socially bounded. Specifically, examination of the behavioural sequence showed that intragroup escalatory acts were only followed by intergroup escalation in 2.2 % of occasions. At a similar level, intergroup escalatory acts were met with intragroup escalation in 2.3 % of occasions. This intragroup escalation following intergroup escalation is perhaps behavioural evidence for the findings that group members may be highly aggressive towards their own deviants (see Stott et al., 2001). This implies that escalatory responses were not simply a spillover, or randomly placed, but instead highly socially focused. It further implies that groups largely attempted to manage their deviants with de-escalatory, rather than escalatory, behaviours. Important to note here is that the behavioural categories we employed for coding

individuals' actions may have been too crude at times. Specifically, to adopt the code 'escalatory' for any aggressive act towards another may misrepresent whether an aggressive response is necessarily 'escalatory' from the perspective of the actor. For example, if an actor responds aggressively towards another, this action may be intended as de-escalatory, to stop further potential acts of aggression by the other. Although speculative, this would be an example of utilising aggression for the purpose of de-escalation, and this would arguably not be an 'escalatory' action. We coded our data with a limited number of broad categories for simplicity, increased statistical power and to follow the previous work of Levine et al. (2011). While we are fairly confident that the 'escalation' code is the correct rule for the majority of aggressive acts, it is also apparent that the coding scheme at present is not sophisticated enough to encompass these exceptions. With a larger dataset and a more fine-grained coding scheme, these behavioural exceptions (for example escalation for the purpose of de-escalation) may be interesting avenues of research to further examine the diversity of strategies employed to achieve conflict resolution.

There are further limitations to our study. Although we controlled for the type of conflict a video depicted, the actions of an outgroup member during an intergroup fight versus an intragroup fight and the actions of an ingroup member during an intragroup fight versus an intergroup fight may vary qualitatively. For example, if a conflict depicts an intragroup fight, we might expect more of the ingroup-directed actions to be hostile (e.g., 'we're having a fight') in comparison to the intragroup actions that occur during an intergroup altercation (e.g., 'we're fighting them, and I'm interacting with one of us'). In the same vein, if a conflict depicts an intragroup fight, we might expect that most of the intergroup actions represent an external parties' attempts to mediate the situation. In this example, it would not be an intergroup action of 'us versus them', but rather an action of 'it's them versus them, and I'm entering their conflict'. These nuances highlight the complexity of the data. Furthermore, by coding both intragroup actions and intergroup actions within the same video, and treating

these coded actions as separate cases for analysis, there may be issues regarding interdependence of the intergroup/intragroup data. For example, one might suspect that any intergroup action (for example, an escalatory push of an outgroup member) may determine, to some degree, the chances of a subsequent intragroup action (for example, a succeeding de-escalatory attempt made by an ingroup member towards their own aggressor). The same argument may be made regarding interdependence of actions more generally. Specifically, the first action (e.g., a de-escalatory act), may make a specific succeeding action more likely (e.g., a further de-escalatory act)—as supported in our sequential analyses. As independence of data is not guaranteed, the results of our analyses must, therefore, be interpreted with caution.

Similar to the research of Levine et al. (2011), this study examined the relationship between the number of active bystanders present and number of escalatory and de-escalatory acts. However, unlike Levine et al.'s (2011) study, cases were divided into videos depicting intragroup fights and videos depicting intergroup fights. Both intragroup and intergroup fights showed a positive relationship between the number of active bystanders and number of de-escalatory acts when controlling for video length. This positive association is perhaps unsurprising given that de-escalatory acts were the most common behaviour, and active bystanders largely contributed to these acts.

There was a non-significant association recorded between the number of active bystanders and number of escalatory acts for intergroup fights. Tentatively, this non-significant finding may be evidence that increased audience presence does not necessarily result in peer group excitation, greater concerns over impression management, partisan actions and increased hostility (Felson, 1982; Jankowski, 1991; W. B. Sanders, 1994; Tomsen, 1997; Weenink, 2014). Interestingly, there was a moderately large negative association between the number of active bystanders and number of escalatory acts for intragroup fights. Although we cannot make any definitive claims about causation (e.g., fights with fewer escalatory acts attract a larger audience), it is possible that the presence and actions of ingroup

members are important for creating the conditions necessary for within-group social regulation and violence reduction. More specifically, ingroup members may explicitly help curtail ingroup conflicts through the direct surveillance and supervision of potential fighters, and also through actions and interventions that signal that certain behaviours are unacceptable to the group (Groff, 2014).

An important finding is that third-parties were present in incidents of public space violence (Planty, 2002). In fact, at least one active bystander was present in each of the 43 videos we analysed. Furthermore, third-parties performed the majority (66.3 %) of actions coded in our material. Over three quarters of these actions were de-escalatory in nature. These third-parties therefore, were more than a generic feature of the ambient context; they were active players who contributed actions, who largely tried to de-escalate (though also escalate) conflicts, and who were found, at the very least, to have some correlational impact on events. Whether the actions of third-parties produced escalation or de-escalation, to an extent, seems to be qualified by whether the aggression happened in intergroup or intragroup contexts.

With these considerations in mind, a more dynamic model of aggression and violence should look beyond the aggressive dyad, and allow for the interaction between all relevant actors in the emerging aggressive episode. More specifically, an improved model of aggression and violence would look at the relative contribution of the context, and people in the context, as much as the dyadic individuals themselves. An improved model of aggression would consider social group categorizations as explanatory components of violence emergence and its regulation. Of particular interest therefore, is the examination of the more established accounts of aggression and violence (GAM and I³ theory) against an alternative model that also incorporates further dynamics between actors who share social groups, collective norms and collective inhibition. Given the ethical and practical challenges of comparing models of aggression dynamically in experimental settings, one viable alternative could be to conduct such works in an agent-based modelling environment.

3 Social Group Formation and Prototypical Positions in an Agent-Based Modelling Environment

Up until this point, we have been largely dissatisfied with current models of aggression that tend to neglect social groups and social context as important determinants of violence, and which are insufficient to explore the dynamics of violence escalation beyond the aggressive dyad. Therefore, over the next two chapters, we aim to test the principles of the general aggression model and the I³ theory in an agent-based modelling environment. We develop an alternative approach that amalgamates the principles from these established accounts with recent computational developments in self-categorisation theory (SCT). We model the dynamics of group formation—exploring how social context can produce escalation to violence—and inhibition of violence.

In this chapter, we begin construction of our agent-based model (ABM) of aggression and violence. The first step of model development is to implement the SCT of metacontrast into an agent-based modelling environment. This metacontrast mechanism will be later used by our agents (in our broader ABM of violence) to form social groups and to derive prototypical positions on violence that are contextually dependent.

This chapter commences with a brief description of what agent-based modelling is and how the methodology has been implemented in social psychology. Our computational work then translates a pre-existing metacontrast prototypicality function—formulated by Salzarulo (2004, 2006)—into our own agent-based modelling environment. In Study 2, we validate our prototypicality function against the existing function of Salzarulo (2004, 2006) that was implemented in an alternative, unspecified Java applet. We conduct replica experiments of Salzarulo's (2004) original work to demonstrate the robustness and validity of our adapted prototypicality function.

In a novel experiment (Study 3), we use the principle of metacontrast to form context-dependent social groups who move together in a physical space. We confirm whether the

social group formation found is in line with the predictions of SCT; in particular, we look at whether agents form social groups with prototypical agents, and whether agents can migrate to different groups depending on the shifting social context.

3.1 Agent-Based Modelling

Agent-based modelling is a computational method that simulates complex and dynamic social systems comprising interacting agents. More specifically, ABMs provide a platform in which a virtual population of autonomous agents iteratively execute a variety of theoretically-derived behaviours. These behaviours—broken down into computational algorithms, condition–action rules and decisions—in turn influence how an agent observes, learns and makes decisions within their world. The continual updating of an agent’s internal state to the changing context subsequently alters future agent–agent, agent–environment and environment–agent interactions. Of particular interest is how these local engagements at a decentralised micro level aggregate over time into emergent and often unexpected higher-level phenomena (Epstein, 1999).

A simple, and classic, example of an emergent higher-level phenomenon is found in Shelling’s (1971) model of racial segregation for US cities. In this model, individual agents were randomly placed on a single cell of a lattice neighbourhood and assigned to one of two differently-coloured groups (the most recent examples being red or green; Wilensky, 1999). In addition to having a ‘colour identity’, agents were assigned a tolerance value that symbolised a mild preference for living around their ingroup. This preference—set at a 3:10 ratio—meant that agents were satisfied to remain in their current location as long as 70 % (7:10) of their local neighbours were of the same colour group. At each iteration, agents were automated to survey their immediate neighbourhood (the eight surrounding lattice cells) and calculate whether their tolerance threshold was satisfied. If the total number of outgroup members exceeded 30 % (3:10), agents would move to a random empty cell on the lattice. Highly replicable results show that after only a few iterations, clearly segregated neighbourhoods

formed (see Figure 3.1). These findings, which would not be expected based on the ‘moderately’ tolerant (according to Schelling, 1971) attitudes of the individual agents, highlight one of the key strengths of ABMs—the ability to see emergent behaviour far greater than the sum of its parts.

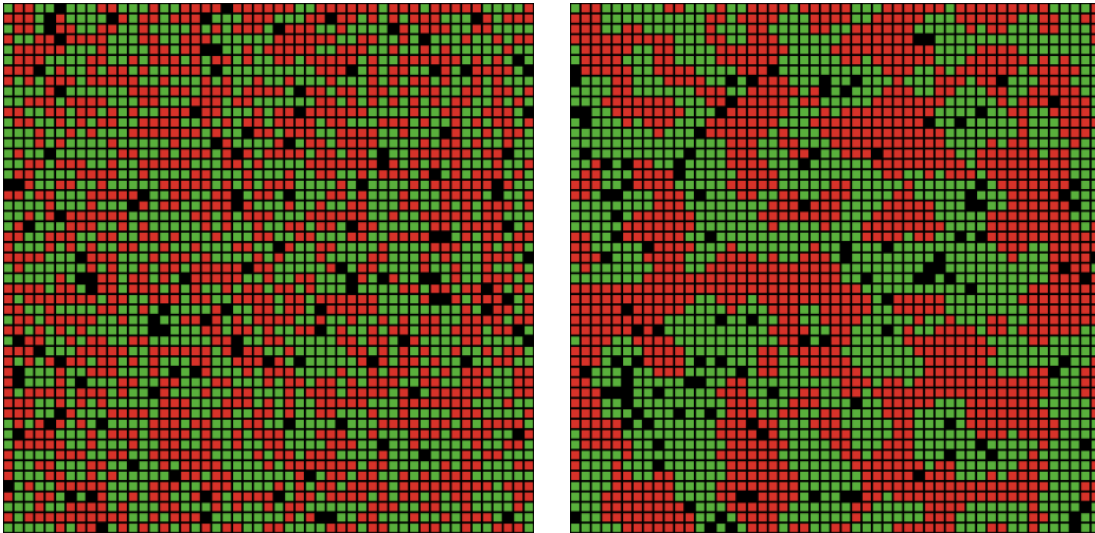


Figure 3.1. Model of Segregation (Schelling, 1971; as operated by Wilensky, 1999)

Other than a dynamic advantage, ABMs hold several other strengths (for a rounded assessment, see Birks & Elffers, 2014). As all agent behaviours need analysing and subsequent formalising from the theoretical level, experimenters are required to be thorough and explicit in these formulations. This prevents theoretical inconsistencies that may have been overlooked by traditional verbal models (Galán et al., 2009). Once programmed, an ABM can generate an infinite amount of simulations, all of which—unlike more traditional experiments—are perfectly controlled and free from external influencing factors. Further, each simulation is saved in its entirety and can be returned to at any point. This allows the experimenter to track, examine and scrutinise any single micro-interaction occurring throughout the simulation’s lifespan. Additionally, the parameters of any model or single simulation may be adjusted with ease to isolate specific effects or to test new hypotheses. These ABM parameters and interactions can be stochastic, thus allowing for randomness between simulations. This allows researchers to ascertain when and how experimental results alter through the inclusion of inherent randomness. With such absolute observational

control—and without the monetary and ethical restrictions of traditional experimentation—the researcher is only limited by their level of expertise, and the constraints of time and computational power (Birks & Elffers, 2014).

Although this advocates the use of ABMs as important instruments of research, it is important to emphasise that ABMs are only models. Congruence between a simulation and the phenomenon of interest is not necessarily evidence that the model reflects reality, or that the processes implemented truly underpin this phenomenon (Epstein, 2006). Further, while a researcher can have an intricate understanding of the processes and parameters programmed into a model, the processes and results that emerge from simulation runs may be opaque, difficult to interpret, uncertain and only moderately reproducible (Manzo, 2014). This means that models tend to include the minimal number of parameters and processes, which restrict both the model's complexity and its predictive power (Leombruni & Richiardi, 2005).

3.2 Agent-Based Modelling in Social Psychology

Although more often utilised in the 'harder' sciences, ABMs are becoming recognised in social psychology. More recently, ABMs have been used to examine opinion dynamics (Salzarulo, 2006), conformity (Iannaccone & Makowsky, 2007), social coordination (Nowak, 2017) group optimal distinctiveness (Smaldino, Pickett, Sherman, & Schank, 2012), segregation in social networks (Henry, Prałat, & Zhang, 2011), social influence across extended social networks (van Rooy, 2012), intergroup destigmatisation (Heinke, Carslaw, & Christian, 2013) and the role that social identity plays in prosocial behaviour during emergency situations (von Sivers, Templeton, Köster, Drury, & Philippides, 2014).

Rather than supplanting the more traditional analytical techniques, ABMs work to complement current methodology. To date, much of social psychology has depended on cross-sectional studies that present isolated and static snapshots of the phenomenon of interest (for methodological critique, see McGrath, Arrow, & Berdahl, 2000). However, by definition, group dynamics are 'dynamic', with complex social phenomena arising from the recurrent

interactions of numerous individuals over an extended period (Postmes et al., 2005; E. R. Smith & Conrey, 2007). While longitudinal studies are more temporally sensitive, they are unable to capture precise changes in behaviour stemming from complex real-time interactions and feedback loops (Smaldino, Calanchini, & Pickett, 2015). ABMs can compensate for these shortcomings and can accommodate feedback loops between all actors and at different levels. This allows the researcher to pinpoint exactly when and how cascading effects occur.

In addition to offering greater dynamism, ABMs are useful instruments in theory development and theory testing. One example that highlights both of these advantages can be seen in the work of Smaldino et al. (2012). Taking Brewer's (1991, 1993) theory of optimal distinctiveness—the principle that individuals seek membership of groups that are moderately sized to best minimise isolation while best maximising differentiation—Smaldino et al. (2012) constructed a computational representation of these social processes in play. While impressed with the accumulated evidence supporting optimal distinctiveness theory (as evident in a review by Leonardelli, Pickett, & Brewer, 2010), Smaldino et al. (2012) recognised methodological limitations in the way that this research was conducted on single participants responding to a static environment. Individuals alter their social identities in response to the changing social context that in turn influences others. Consequently, as individuals shift between groups, the levels of distinctiveness and attractiveness of each group update. This is a continual and fluid process and difficult to capture using lab-based experimentation.

Akin to the work of Schelling (1971), Smaldino et al. (2012) randomly placed individual agents on a single cell of a lattice neighbourhood and assigned two simplified variables: membership to a group (social identity) and a preference ratio for one's own group to be distinct from others (degree of optimal distinctiveness). At each iteration, agents were asked to survey their local neighbourhood and assess the numbers of agents in each social identity. If the agent was to find another group closer to its optimal level of distinctiveness,

the agent would switch its social identity and join this other group. This, in turn, would influence the succeeding agents in their calculations, thus aggregating micro-behaviour into emergent macro-behaviour. After several iterations, results showed that stable and optimally-distinct groups would form—all of which contained agents with a satisfied degree of optimal distinctiveness.

The implications of these simulations were twofold. First, the model was able to circumvent many of the issues inherent in static laboratory testing. With relative ease, the model was able to represent optimal distinctiveness as something both fluid and influencing—an assertion that could only be made putatively up to that point. Second, the model uncovered unexpected additional results that advanced the field's knowledge of optimal distinctiveness theory. More specifically, while optimal groups formed in scenarios where agents were attentive to their moderately-sized local neighbourhoods (i.e., they knew the identities of their local neighbours only), this effect disappeared as neighbourhood parameters expanded to a global level (i.e., where agents knew the identities of all other 2499 agents). In this enlarged context, groups became unsettled and overly large with a suboptimal degree of distinctiveness. Alternatively, with severely restricted neighbourhoods, (i.e., where agents knew only the identities of their immediate neighbours), groups that formed were homogeneous and assimilative. These computational findings not only support the assumption that individuals are most satisfied in a moderately-sized group, a central tenet of optimal distinctiveness theory, but also suggest that moderately-sized networks are the most accommodating to facilitate the emergence of optimally distinct groups. This helps researchers to better understand the role and restrictions of social networks—both locally and globally—in group formation (something not explicitly considered in traditional optimal distinctiveness literature, see Smaldino et al., 2015).

In sum, and as the example above alludes to, ABMs are useful instruments for the social psychologist. The formulation of an ABM forces the experimenter to thoroughly

understand and explicitly state the theoretically-driven mechanisms involved in a phenomenon. This approach removes illogical inconsistencies that are commonplace in traditional verbal models. Once constructed, ABMs can improve theoretical purchase and complement current methods by further replicating laboratory findings in a dynamic and fluid manner. As a result, ABMs can help inform and guide new directions for research. This involves fewer ethical and monetary restrictions, and allows the experimenter absolute observational control.

3.3 Study 2: Metacontrast and Social Group Formation in an Agent-Based Modelling Environment

3.3.1 The Model: Purpose, State Variables and Scales

The model description follows the standard overview, design concepts and details (ODD) protocol for describing individual models and ABMs (Grimm et al., 2006, 2010). The model was implemented in NetLogo v. 5.2.0 (Wilensky, 1999).

Purpose. The main purpose of this study is to model social group formation and prototypicality in an agent-based modelling environment. In SCT (Turner et al., 1987), social groups are represented as prototypes. These prototypes are fuzzy sets of interrelated attributes (attitudes, feelings, opinions and perceptions) that capture the similarity within a group and accentuate the differences of those not within the group, thus discriminating the relevant category from its outgroups (Hogg, 2006; Hogg & Reid, 2006). More intricately, prototypes are determined through *metacontrast*, which is the principle that:

Any collection of stimuli is more likely to be categorised as an entity (i.e., grouped as identical) to the degree that the differences between those stimuli on relevant dimensions of comparison (intra-class differences) are perceived as less than the differences between that collection and other stimuli (inter-class differences) (Turner et al., 1987, p. 47).

A prototype sets the boundaries surrounding an identity and determines who may be considered ingroup or outgroup; however, these boundaries are not fixed. Rather, categories can be defined narrowly and exclusively or broadly and inclusively depending upon the specific context (Haslam, Reicher, & Levine, 2012; Oakes, Haslam, & Turner, 1994).

Seeking to differentiate one's own group from others, and in line with metacontrast, this relevant prototype is rarely the central position of the group, or even a tangible individual. Rather, the prototype is a hypothetical entity that optimises distinction from other outgroups, while also best exemplifying the consensus of the ingroup. This typically means that the prototype is more extreme than the central position of the group, pushed away from the average by intergroup comparisons and outgroup aversion (Haslam & Turner, 1992; McGarty, Turner, Hogg, David, & Wetherell, 1992). Further, this prototype and the category boundaries are highly fluid and shift as the context and relevant social comparison alter (Hogg, 2004; Hogg & Terry, 2000).

Beyond setting the inclusion boundaries of a social group (who is ingroup and who is outgroup), a prototype also describes and prescribes how a member ought to behave (Hogg & Reid, 2006). "In this sense prototypes are norms; that is, because a particular perception, behaviour or attitude is shared within a group, it is normative of that particular group ... thus, prototype-based attitudes are normative" (J. R. Smith & Hogg, 2008, p. 344). Through categorisation and identification with a group containing the self (ingroup), individuals depersonalise themselves and subsume these prototypically relevant attributes (Turner et al., 1987). This assimilation leads to individual behaviour congruous with that of the group, with adherence to collective norms, values and expectations. As the prototype is typically more extreme than the mean of the combined members, this conformity towards a norm, value or expectation may represent polarisation—a phenomenon stronger in magnitude in relation to the extremity of the comparative social context (McGarty et al., 1992).

In brief, individuals cognitively represent social groups as prototypes. A prototype is a context-dependent representative entity that optimises distinction from outgroups, while also best exemplifying the consensus of the ingroup. When categorised and depersonalised, the position of this fluid prototype holds significant sway in determining which behaviours, values and positions are default and deemed acceptable. Assimilation to the prototype in more relevant intergroup comparisons can lead to repulsion from other groups (*polarisation*), which may explain extreme actions, behaviours and beliefs beyond the sum of the group.

The main purpose of this model, therefore, is to translate the social-psychological principle of metacontrast (Turner et al., 1987) into an agent-based modelling environment. We validate our model against the existing metacontrast model of Salzarulo (2004, 2006) that was not implemented in NetLogo, but in an alternative, unspecified Java applet. Once tested and verified, this metacontrast model will be used to create context-dependent social groups that are physically mobile (Study 3). This metacontrast model will ultimately be implemented into a broader social model of aggression and violence (Studies 4, 5a and 5b).

State variables and scales. The main entity of this model is the agent. The agent represents an individual who is characterised by an interrelated attribute they hold that can be used for metacontrast comparisons with others. For our simulation, this comparative attribute (opinion, attribute, feeling or position) is an opinion ranging from 0.00 to 1.00. The simulation runs on a square lattice of 25 x 25 cells. Each agent has an identity number to help differentiate it from others. Each agent also has a neighbourhood (\mathcal{N}_i) that is the radius of the local vicinity in which it interacts with others (see Figure 3.2). This neighbourhood represents a physical parameter (or vicinity) that surrounds the agent, and can be expanded or restricted accordingly. For this specific simulation, the \mathcal{N}_i radius is set exceptionally high at 26. The purpose of this large neighbourhood size is to engulf the entire 25 x 25 grid and ensure a full-mixing condition in which each agent can interact with all others, without requiring

movement. As such, each individual agent is immobile and remains on their assigned cell for the entirety of the simulation (this changes in Studies 3–5).

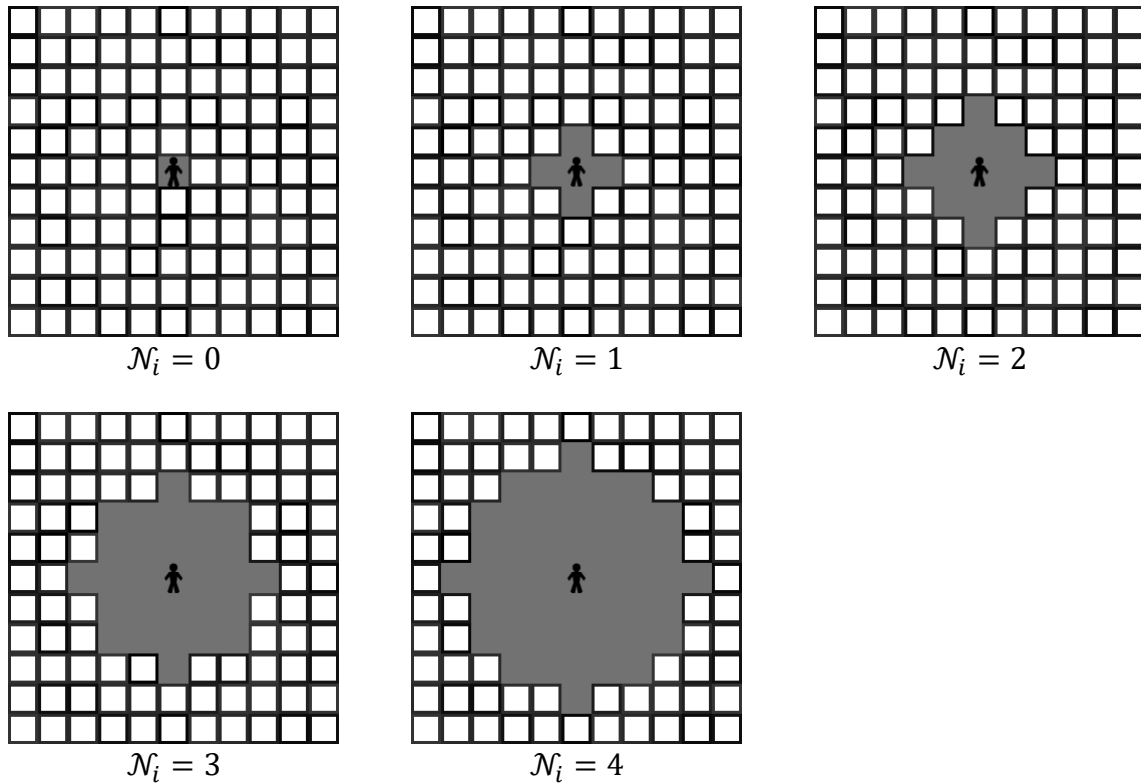


Figure 3.2. Example of NetLogo in-radius neighbourhood (grid size: 11 x 11)

3.3.2 Process Overview and Scheduling

The model has one main process: to run the prototypicality function as described by Salzarulo (2004). This prototypicality function considers the opinions of all agents within the neighbourhood and determines how prototypical any supplied opinion value (ranging from 0.00 to 1.00) is of that local context. This is calculated via a metacontrast calculation, with each opinion a point of comparison. Following Salzarulo (2004, 2006), we define a function that determines how prototypical any opinion value (a) is within the local neighbourhood as the weighted difference of the distance between the ingroup and outgroup opinions:

$$P(a, \mathcal{N}_i) = \gamma \times d_{inter}(a, \mathcal{N}_i) - (1 - \gamma) \times d_{intra}(a, \mathcal{N}_i)$$

where $\gamma \in [0,1]$ is a parameter signifying the outgroup aversion (a value of zero equals no repulsive effect due to the outgroup).

We can find the mean distance between any inputted opinion value (a) and the opinions of the ingroup neighbours using the d_{intra} function:

$$d_{intra}(a, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} m(a, a_j) \|a - a_j\|^2}{\sum_{j \in \mathcal{N}_i} m(a, a_j)}$$

where $m(a, a_i) = \exp(-\frac{\|a - a_i\|^2}{\omega^2})$ is a ‘fuzzy membership function’ that determines how similar two opinions are to one another, and $\omega \in [0,1]$ is a width parameter determining the tolerance of differing opinions (a ω value close to zero signifies that very few different opinions are considered ingroup opinions; a value of one signifies that all opinions, regardless of their diversity, are considered ingroup).

Likewise, we can determine the mean distance to the outgroup neighbours’ opinions using the d_{inter} function:

$$d_{inter}(a, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} (1 - m(a, a_j)) \|a - a_j\|^2}{\sum_{j \in \mathcal{N}_i} (1 - m(a, a_j))}$$

The prototypical opinion a^* for any agent is found by first locating the local maximum (highest possible return value) of $P(a, \mathcal{N}_i)$. This is achieved by feeding a dummy opinion (a) (originally set at zero and adjusted in 0.001 increments up to one) into the $P(a, \mathcal{N}_i)$ function. Each increment returns a P score, which is a score of how prototypical this inputted value is, when calculating a metacontrast ratio of all agents within the neighbourhood. Essentially, the higher the P score returned, the closer this inputted opinion is to being prototypical. The P scores generated from each increment are best visually plotted in a scatter plot (with opinions from 0.00 to 1.00 on the X-axis and P score on the Y-axis, see Figures 3.4 and 3.6). Plotting these variables results in a hilly terrain. Those agents’ opinions that fall within a hill (visually represented as falling under the same hill-shaped umbrella) share a prototype category. The local maximum of an agent’s hill represents the optimal prototypical opinion for that agent. Each agent adopts the opinion of its prototype. Therefore, the prototypical position for any

agent may not be that agent's original opinion. Rather, it is the opinion that the agent perceives as prototypical for its category.

3.3.3 Design Concepts

Emergence. Emergent phenomena, such as consensus of opinions, polarisation of opinions and extremism, may arise as an emergent property of intergroup comparisons and outgroup aversion.

Adaptation. The prototypicality curves generated via the function will determine the prototypical opinion for each agent and their category. This prototypical opinion is context dependent and will alter as the comparative context shifts. This means that agents who were once considered outgroup members may be recategorised as ingroup members if the introduction of a more extreme opinion presents itself. After the prototypical boundaries are determined, the agent will alter its opinion to be congruous with its optimal prototypical position. However, as a prototype is typically more extreme than the central position of the group (having been pushed away from the average position by intergroup comparisons and outgroup aversion; Haslam & Turner, 1992), there is a possibility that an agent's opinion will become more extreme than its original position. This may result in polarisation of opinion—a phenomenon stronger in magnitude in relation to the extremity of the comparative social context (McGarty et al., 1992).

Objectives. The objective of this simulation is to run the prototypicality function, determine the prototypical maxima and have agents adopt the most prototypical position of its category.

Sensing. The function recognises the opinions of all agents within the local neighbourhood that, in this example, encompasses the entire grid.

Interaction. For this first study, the agents themselves do not interact. Rather, the prototypicality function will be fed varying opinions in small increments. The prototypicality

function will compare this fed opinion to the opinions of all agents upon the grid using a metacontrast calculation. It will determine how optimally prototypical *any* opinion is from 0.00 to 1.00. The function will determine the optimally-prototypical position for any agent within the simulation.

Stochasticity. The sequential order in which the function considers the inputs, the opinions of the relevant agents, is from smallest value to largest.

Observation. Each incremental opinion value (from 0 to 1 in increases of 0.001) is fed into the prototypicality function. Each incremental opinion value and its corresponding P value is plotted as graphical outputs (see Figures 3.4 and 3.6).

3.3.4 Initialisation, Inputs and Submodels

This simulation was initialised with 12 agents (experimental scenario A) and six more agents were then later added (experimental scenario B). Each agent was created with a pre-assigned opinion (more detail is provided below). We set the width (ω) and aversion (γ) parameters at 0.36 and 0.08⁶ respectively, as calibrated by Salzarulo (2004) using the experimental data of (Haslam & Turner, 1995) in which individuals assimilate with, or polarise from, others based on their relative positions on a pragmatic–idealist scale.

The model does not include any external input data driving environmental factors. The model also does not include any submodels.

3.3.5 Simulation Experiments

This simple simulation experiment was conducted to demonstrate the robustness and validity of our prototypicality function against the original formulation and data provided by Salzarulo (2004). The experiment contained two hypothetical scenarios (scenario A and

⁶ Salzarulo's (2004) original parameters calibrated against the experimental data of Haslam and Turner (1995) were $\omega = 0.3556$ and $\gamma = 0.08096$ respectively.

scenario B). These were designed to alter comparative context between agents and to demonstrate this impact on prototypical categorisation.

3.3.6 Descriptions and Results of Scenarios

Experimental scenario A. Imagine, in an adapted scenario from Oakes (1996), that there are 12 individuals: six of whom are physicists and six of whom are biologists. These individuals are quizzed regarding the degree to which they consider their work to be quantitative or qualitative (on a scale from 0.00 to 1.00). Let us assume, as Salzarulo (2004) does, that each of the six physicists consider themselves quantitative researchers, with positions of {0.05, 0.06, 0.07, 0.08, 0.09, 0.10} respectively. Let us also assume that each of the six biologists also have a quantitative position, but that each position is more qualitative than their physicist counterparts {0.40, 0.41, 0.42, 0.43, 0.44, 0.45} respectively. In a scenario where the physicists are by themselves and no comparison is made, Salzarulo (2004) assumes that the prototypical position would be the mean (centre) of the group (0.075). However, as the comparative context changes and we introduce the biologists, this position should vary as a function of minimising the intraclass differences and maximising the interclass differences (Oakes, 1996).

To test this, we introduce these 12 pre-programmed individual agents into our agent-based modelling environment. Assume that each of the 12 individual agents has no information about its prototypical position or of any potential shared categories. Feed an opinion value (a) in 0.001 increments into the $P(a, \mathcal{N}_i)$ prototypicality function (from 0.00 to 1.00). For demonstrational purposes, plot each incremental value (a) onto an X-axis, with its corresponding P score plotted onto the Y-axis. Ask each agent to adopt the corresponding opinion value of its local maximum (the local peak of the prototypical category under which an agent's original position is situated).

Results of experimental scenario A. The first graph depicts the data obtained by Salzarulo in his 2004 simulation with an unspecified Java applet (see Figure 3.3). The original opinions of each of the individuals are indicated by the small vertical bars (whiskers) above the X-axis. The function has returned two humps which indicate that two groups exist. The prototypical group boundaries are represented by these humps. One group contains the opinions of all of the physicists and the other group contains the opinions of the biologists. The prototypical positions of each group correspond to the local maxima of their hump, and are indicated graphically by a black triangle.

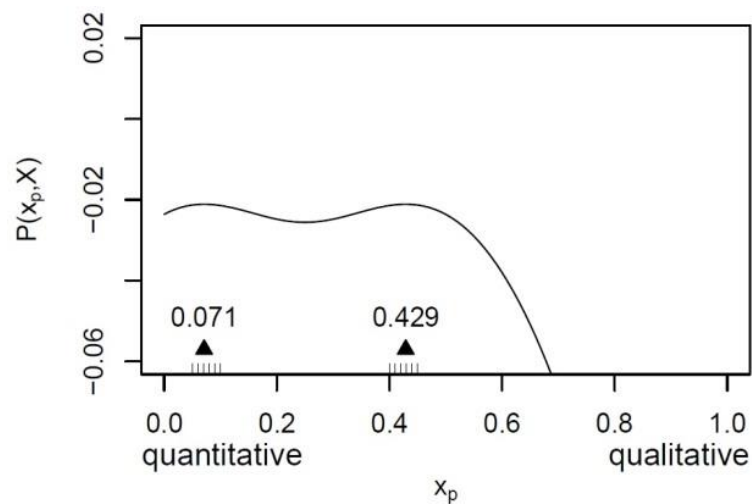


Figure 3.3. Prototypicality curves for physicists and biologists (taken from Salzuro, 2004; Salzarulo's x_p is equivalent to our a)

The prototypical opinion of 0.071 suggests that in the presence of the biologists, the physicists have slightly polarised their opinion (0.071) away from their original mean central position (0.075), and thus away from the outgroup. The biologists also have marginally polarised their opinion away from their central position (0.425) and now have a new opinion of 0.429. This indicates that in the presence of the biologists, the physicists feel that their work is slightly more quantitative. The opposite is true for the biologists.

The second graph depicts the data obtained by us, the experimenters, when running our replica simulation in NetLogo (see Figure 3.4). The graph is identical to Figure 3.3 with

two formed groups, each with slightly polarised opinions as represented by the vertical bar lines (physicists: 0.071; biologists: 0.429).

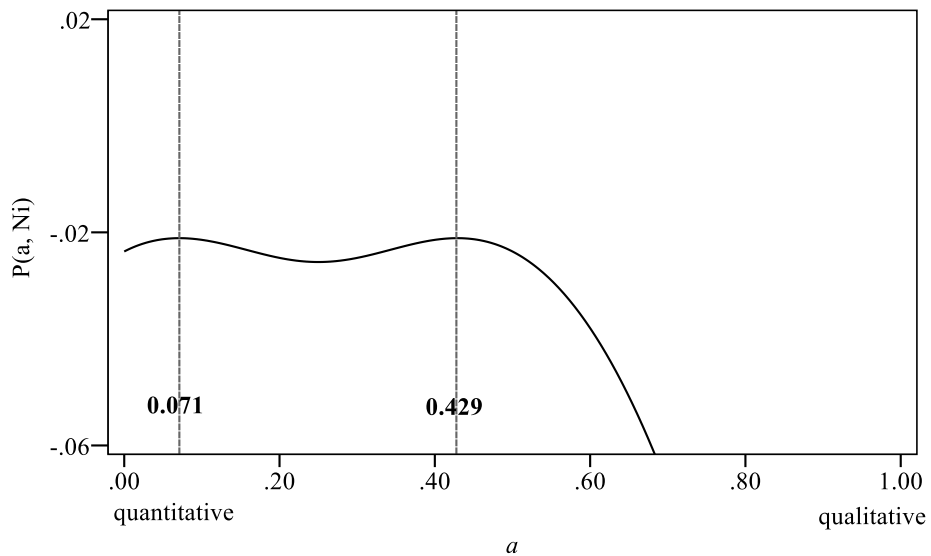


Figure 3.4. Investigators' prototypicality curves for physicists and biologists

Experimental scenario B. Imagine a new scenario in which we retain the physicists and biologists and reset their original positions, but add a third set of six individuals who are social scientists. Similar to Salzarulo (2004), give each social scientist a highly qualitative position $\{0.85, 0.86, 0.87, 0.88, 0.89, 0.90\}$, thus enlarging the comparative context, and introduce the 18 pre-programmed individual agents into the agent-based modelling environment. Once again, assume that each of the individual agents has no information about its prototypical position or of any potential shared categories. Feed an opinion value (a) in 0.001 increments from 0 to 1 into the $P(a, \mathcal{N}_i)$ (prototypicality) function and replot the corresponding P scores.

Results of experimental scenario B. The following graph depicts the data obtained by Salzarulo in his 2004 simulation with an unspecified Java applet (see Figure 3.5).

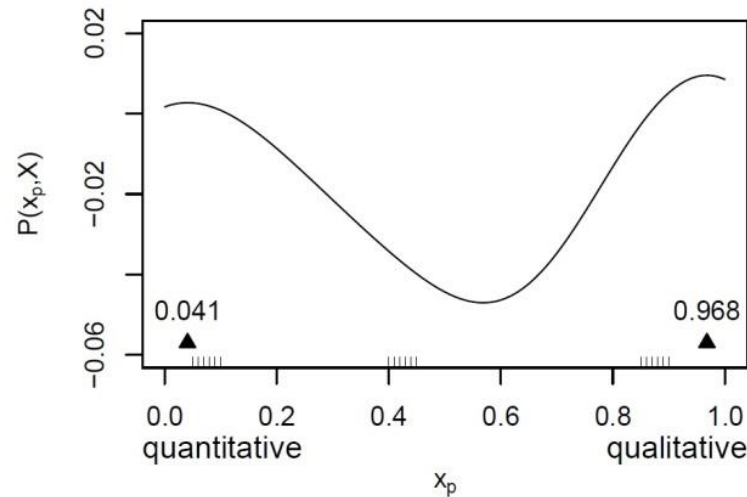


Figure 3.5. Prototypicality curves for physicists, biologists and social scientists (taken from Salzuro, 2004; Salzarulo's x_p is equivalent to our a)

Again, the individuals' original opinions are represented by the small vertical bars above the X-axis and the prototypical group boundaries are defined by the humps. The two humps signify that a total of two groups have formed. One group contains both the physicists and biologists, who now exist as a shared category in the presence of the more extreme social scientists. The social scientists exist as the second group. The prototypical opinion of 0.041 suggests that in the presence of the social scientists, the physicists and biologists have polarised their opinions away from their original mean central positions (0.075 and 0.425 respectively). The social scientists, with a prototypical opinion of 0.968, are also polarised away from their original opinion of 0.875. This means that in the presence of the social scientists, the physicists and biologists feel that their work is more quantitative. The opposite is true for the social scientists.

Figure 3.6 depicts the data obtained by us, the experimenters, when running our corresponding simulation in NetLogo and is identical to Figure 3.5, with two groups forming: one containing the physicists and biologists, and the other containing the social scientists. Once again, the opinions of both groups have been polarised away from their original

positions, resulting in prototypical opinions of 0.041 (physicists and biologists) and 0.968 (social scientists).

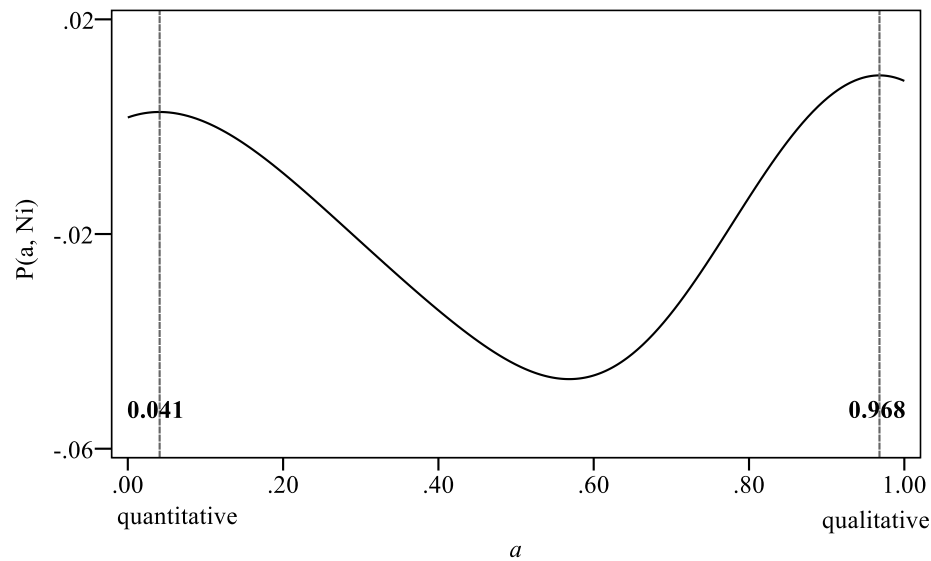


Figure 3.6. Investigators' prototypicality curves for physicists and biologists, and social scientists

3.3.7 Discussion

The objective of our study was to translate the social-psychological principle of metacontrast into an agent-based modelling environment, and to validate the metacontrast prototypicality function against the original formulation from which it was derived (Salzarulo, 2004, 2006). Similar to Salzarulo (2004), we created two theoretical and comparative scenarios that allowed social group categorisation, polarisation and superordinate recategorisation. After running these two scenarios with our NetLogo-adopted metacontrast function, our results (in both scenarios) were identical to Salzarulo's (2004) original findings. This is reassuring, as we operationalised our own formulations based on his work in an alternative programming platform (which used alternative pseudo-random number generation and default digit precisions in numbers). Given the identical values of all prototypical opinions, and the qualitatively equivalent behaviours produced, we may conclude that our model and Salzarulo's (2002) original model are a good fit.

Theoretically, our first experiment provides some interesting insights. Akin to SCT (Turner et al., 1987), we found that relatively stable groups formed under the principle of metacontrast. These groups were formed from individual opinions on a topic with no prior information given about any potential categorisations. The initial population of agents—comprising physicists and biologists—produced two prototype groups. These prototypes captured the similarities within each group and accentuated the differences of those not within their group. More interestingly, each agent was correctly categorised based on occupation and each individual reached consensus with other members within their occupational group, while also polarising away from the relevant outgroup. Such behaviour is in line with the predictions of SCT.

Also interesting is what happened after the introduction of the social scientists. This change in the comparative context led to the creation of a more inclusive superordinate category: ‘hard scientists’. In this category, the physicists and the biologists now perceived their work as extremely quantitative. Outside this category, the social scientists adopted a far more qualitative position. These collective opinions were found to be more extreme than any of the original positions of the members involved, thus providing computational evidence for how polarisation, and potentially extremism, may emerge from the comparative context. These findings are also faithful to the tenets of SCT (Turner et al., 1987) and the theoretical assumptions outlined by Oakes (1996).

Taken together, our metacontrast prototypicality function, adapted from Salzarulo (2004), is able to create group boundaries and category relevant prototypes. These groups, rather than being explicitly pre-set, are dynamic and form as a result of the comparative context. The model produces emergent phenomena such as consensus and polarisation. Our next objective is to use this metacontrast model to form mobile social groups that will move around together in physical space.

3.4 Study 3: Self-Categorisation, Social Group Formation and Physical Flocking

SCT provides insight into how individuals converge to form social groups. Yet, and in common with most models concerning group dynamics, the empirical findings supporting SCT group formation are primarily obtained from small groups of participants in cross-sectional studies. However, by definition, group dynamics are ‘dynamic’, with complex social phenomena arising from the recurrent interactions of numerous individuals over an extended period. Given this consideration, we explore the tenets of SCT in a dynamic computational modelling environment. More specifically, in this study (Study 3) we randomly generate 100 individual agents, each with two interrelated attributes (redness and greenness values) that are used to determine social group prototypes. Agents are instructed to determine a social category prototype following the basic SCT principle of metacontrast, and to move together with the individual best representative of this prototypical position. These simple instructions at the micro level result in larger collective social group formation at the macro level. More interestingly, the number of social groups and the average number of agents per social group do not settle over time, but rather fluctuate with the constant changes in local comparative contexts. In line with SCT predictions, our results show that those most likely to switch social groups—‘migrators’—have increased psychological distance from their original prototype in comparison to those who remain—‘remainers’. Physical distance from the centre of the group and group size are not predictive of migration.

3.4.1 Model: Purpose, State Variables and Scales

The model description follows the standard ODD protocol for describing individual models and ABMs (Grimm et al., 2006, 2010). The model was implemented in NetLogo v. 5.2.0 (Wilensky, 1999).

Purpose. To our knowledge, there exists no ABM in which dynamic group membership determines locomotion. Typically, agents in simulations are given a fixed social

group at the pre-set and then move with their group as a result of this pre-assignment. The main purpose of this model is to create social groups that can then move together in physical space via the principle of metacontrast. Akin to SCT, these groups are not fixed but are context dependent. Ultimately, in this simulation, agents are mobile and will determine a directional heading derived from their fellow group members. This leads to dynamic group flocking that can be implemented into a broader social model of aggression and violence (Studies 5–6).

State Variables and Scales. The main entity of this model is the agent. Unlike the previous study (Study 2), the agent is characterised by two opinions that it holds on two separate topics—arbitrarily defined as ‘redness’ and ‘greenness’. These opinions range from 0.00 to 1.00 and are randomly assigned from a uniform distribution. As such, it is possible that one agent has a high opinion of red and a low opinion of green, a high opinion of both, a low opinion of both, or any possible combination ranging from 0.00 to 1.00 on either dimension. The simulation runs on a square lattice of 25 x 25 cells⁷ with a one-cell thick world boundary.⁸ As agents are mobile, each agent has an X-coordinate and Y-coordinate that denotes its location on the physical grid. Each agent also has a heading that determines the direction it will take with the next physical step. As with Study 2, each agent has an identity number to differentiate it from another, and a local neighbourhood (\mathcal{N}_i) that is the local vicinity in which that agent interacts with others. Unlike Study 2, this neighbourhood is restricted ($\mathcal{N}_i = 3$) and thus allows interactions only with those agents within a smaller vicinity (see Figure 3.2). The model runs in iterations or ‘ticks’ in the NetLogo language. In a single iteration, each agent produces a prototypicality calculation, links with one agent best representative of this prototype, adopts one physical heading and takes one step forward.

⁷ Although a square lattice of 25 x 25 cells is somewhat arbitrary, no differences in the main results of any of the ABM studies were found in smaller (16 x 16 cells) or larger (32 x 32 or 43 x 43 cells) world alternatives.

⁸ The thick world boundary caused agents to bounce back off the outer-perimeter (like light reflecting off a mirror), ensuring more mixing between agents.

3.4.2 Process Overview and Scheduling

The model has two main processes: i) assigning agents to physical groups via metacontrast, and ii) providing a physical heading for the agent's movement that is a product of this group.

Similar to Study 2, assume that each agent i has opinions about a number of topics⁹— in this simulation: redness (\mathcal{R}) and greenness (\mathcal{G}). For each agent, collect these opinions in an opinion vector ($a_i = (a_i(\mathcal{R}), a_i(\mathcal{G}) \dots)$). Following Salzarulo (2004, 2006) and Study 2, we define a function that determines how prototypical any opinion value (a) is within the neighbourhood. This is the weighted difference of the distance between the ingroup and outgroup opinions:

$$P(a, \mathcal{N}_i) = \gamma \times d_{inter}(a, \mathcal{N}_i) - (1 - \gamma) \times d_{intra}(a, \mathcal{N}_i)$$

where $\gamma \in [0,1]$ is a parameter signifying the outgroup aversion (a value of zero equals no repulsive effect due to the outgroup).

We can find the mean distance between agent i 's opinion and the opinions of the ingroup neighbours using the d_{intra} function:

$$d_{intra}(a, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} m(a, a_j) \|a - a_j\|^2}{\sum_{j \in \mathcal{N}_i} m(a, a_j)}$$

where $m(a, a_i) = \exp(-\frac{\|a - a_i\|^2}{\omega^2})$ is a 'fuzzy membership' function that determines how similar two inputted agents are to one another and $\|a - a_i\|^2$ contains all values of the opinion vector ($((a(\mathcal{R}) - a_i(\mathcal{R}))^2 + (a(\mathcal{G}) - a_i(\mathcal{G}))^2 \dots)$). In accordance with Study 2, $\omega \in [0,1]$ is a width parameter that determines the tolerance of differing opinions (a ω value close to zero signifies that very few different opinions are considered ingroup opinions; a value of one signifies that all opinions, regardless of their diversity, are considered ingroup opinions).

⁹ This can feasibly be increased to any number of the researcher's choice.

Likewise, we can determine the mean distance to the outgroup neighbours' opinions using the d_{inter} function:

$$d_{inter}(a, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} (1 - m(a, a_j)) \|a - a_j\|^2}{\sum_{j \in \mathcal{N}_i} (1 - m(a, a_j))}$$

Once again, the prototypical opinion a^* for agent i is found by locating the local maximum of $P(a, \mathcal{N}_i)$. Unlike Study 2, where prototypes were plotted on a global level by adjustments of 0.001 (0 to 1), here, each agent is required to find its own locally bound prototypical position itself. The prototypical opinion a^* for agent i is found by locating the local maximum (highest possible return value) of $P(a, \mathcal{N}_i)$. This is achieved by slightly adjusting a dummy opinion (originally set as one's opinion) and feeding this new value into the $P(a, \mathcal{N}_i)$ function. So, for example, if an agent's opinions are $a_i = (a_i(\mathcal{R}), a_i(\mathcal{G})) = a_i = (0.2, 0.9)$, the agent would first feed this [0.2, 0.9] vector into the $P(a, \mathcal{N}_i)$ function. This will return a P score. The agent will then randomly adjust this vector value by a marginal amount between 0.009 and 0.011 ($M = 0.01$) or between -0.009 and -0.011 ($M = -0.01$) (e.g., $a_{iDummy} = (0.2096, 0.893)$). This dummy vector value is then fed into the $P(a, \mathcal{N}_i)$ function. If the returned P score is an improvement (higher value) on the previous P score, then this opinion is closer to being prototypical and is retained. This procedure is then repeated until no further improvements on the P score can be made (after 20 attempts with no improvement of P). If an adjusted opinion does not yield an improved P score, then the new opinion is further away from being prototypical and is dismissed. Optimisation of P, while requiring large amounts of calculations, is successful in capturing the prototypical positions of agents (Salzarulo, 2006).

Once all agents have determined their colour prototype, it is time to determine the physical heading that each agent will take for movement. Instruct all agents to survey their neighbourhoods. If they are alone, and have no neighbours to flock with, set a random heading (right-turn random 45 degrees and left-turn random 45 degrees). If the agent is not

alone and there exists one or more other agents in the neighbourhood, then the agent is instructed to create a link with agent k , the agent (excluding self) most similar to the individual's ideal coloured prototype. This connects agents with similar prototypes like a chain, which in turn creates the larger flocking group (see Figure 3.7). Once all agents (who are not alone) have linked to their prototype and all groups are fully connected, each group is assigned a group ID number.

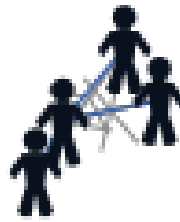


Figure 3.7. An example of connected flocking groups (agents shaded by heading)

After all groups have been assigned, agents are instructed to set their heading the heading of the chained member closest to one's prototypical opinion via the MacLennan (2007) NetLogo flocking code.¹⁰ Once the heading of all agents has been determined, all agents are instructed to take one physical step forward.¹¹ Agents are thus updated once per iteration according to the following steps:

- 1) Find the agents in the neighbourhood \mathcal{N}_i of agent i (including self).
- 2) If there are no agents in the neighbourhood, turn by a small random angle and end turn.
- 3) If there are other agents in the neighbourhood, find the prototypical opinion a^* for agent i , which is the location of the local maximum of $P(a, \mathcal{N}_i)$.

¹⁰ Available from <http://web.eecs.utk.edu/~mclennan/Classes/420-594-F07/NetLogo/Flock.nlogo>

¹¹ `Fd .33` in NetLogo language; deceleration/acceleration is applied so that trailing flock members can catch up (see MacLennan, 2007).

- 4) Identify agent k , the agent (excluding self) with the opinion closest in opinion space to agent i 's prototype a^* .
- 5) Create a link with agent k .
- 6) Share the heading of agent k and end turn.
- 7) Instruct all agents to take one step forward and repeat steps.

3.4.3 Design Concepts

Emergence. Group dynamics emerge from the metacontrast calculations of single agents. With the mobile rules incorporated, individuals will form physical groups that move across the spatial grid together under shared headings.

Adaptation. An agent will determine the prototypical opinion for their category and share headings with those closest to this opinion. This means that agents will tend to cohere around prototypical agents, ignoring the headings of those less prototypical and outgroups with differing opinions. In this simulation, groups—unlike alternative flocking methods—are not fixed. Rather they are context dependent and alter with the comparative situation. This means that agents who were once deemed outgroup members may be recategorised as ingroup members. While one may not expect a huge migration between groups, it is perfectly plausible that a group member may join another group that crosses its path if members of this new group (through the metacontrast calculation) are a better prototypical fit. As agents' knowledge of others is restricted by their immediate neighbourhoods (set at $\mathcal{N}_i = 3$), it is possible that agents will be in suboptimal groups until a more optimal group is encountered.

Objective. The objective of any agent is to find the most prototypical ingroup colour opinion and to flock with those agents within the neighbourhood best representative of this prototype.

Sensing. Agents are assumed to know the opinions of all agents (including themselves) who reside within their neighbourhood.

Interaction. Agents interact by considering their own and others' opinions to determine their category's prototypical opinion. Agents create a physical link with the agent best prototypical of this category.

Stochasticity. The order in which agents calculate their metacalculations is pseudo-random. Agents devise their headings and take one step forward simultaneously.

Observation. All agents' IDs, flocking group IDs, colour opinions, prototype values, derived P values, X-coordinates and Y-coordinates are exported. Each visual output frame is saved as an image that is then amassed as a simulation movie. The NetLogo world is saved in its entirety and can be returned to for further scrutinisation or data abstraction.

3.4.4 Initialisation, Inputs and Submodels

Simulations are initialised with a population of 100 agents. All agents are created with a uniformly-distributed random redness and greenness opinion between 0.00 and 1.00. Each agent is placed on the square 25 x 25 lattice with a random X-Y coordinate starting location and a random heading. The neighbourhood for each agent is set at three to ensure interactions only with those agents who are present locally. We set the aversion (γ) parameter to 0.08 and divided the 0.36 width (ω) parameter by two to 0.18 so as to accommodate for the increase in comparative dimension (from one-dimensional in the previous study—quantitative work—to two-dimensional in this work—redness and greenness; for a discussion of the aversion and width parameters, see Salzarulo, 2006).

The model does not include any external input data driving environmental factors. The model also does not include any submodels.

3.4.5 Simulation Experiment

This simulation experiment was conducted to investigate if the principle of metacontrast could produce dynamic social groups who flock together in physical space. We hypothesise that the groups forming under the steps described above will be statistically more similar in physical space (H1a) and opinion space (H1b) than expected by chance. In line with SCT predictions, we further hypothesise that agents may migrate between social groups (H2) through fluctuations in prototypicality and social context. We further predict that those individuals most likely to migrate from their social groups will have increased psychological distance (H3) from their prototypes than those individuals who remain.

3.4.6 Results

Figure 3.8 depicts the world generated at set-up (iteration zero). The agents appear randomly distributed across the physical grid and do not belong to any physical flocking groups (a total of 100 groups exist with each containing one individual). Figure 3.9 shows the world after 3588 iterations, which is an iteration that has been chosen randomly (between 2500 and 5000 iterations). Agents appear clustered together in groups. Data obtained from the simulation indicate that at iteration 3588, there is a total of 13 groups with a mean of 7.69 group members per group ($SD = 3.73$, min. = 2, max. = 15). A full movie of this flocking simulation up to 5000 iterations can be viewed at <https://youtu.be/EztawaRg8gM>.



Figure 3.8. Simulation at iteration zero (agents shaded by heading)

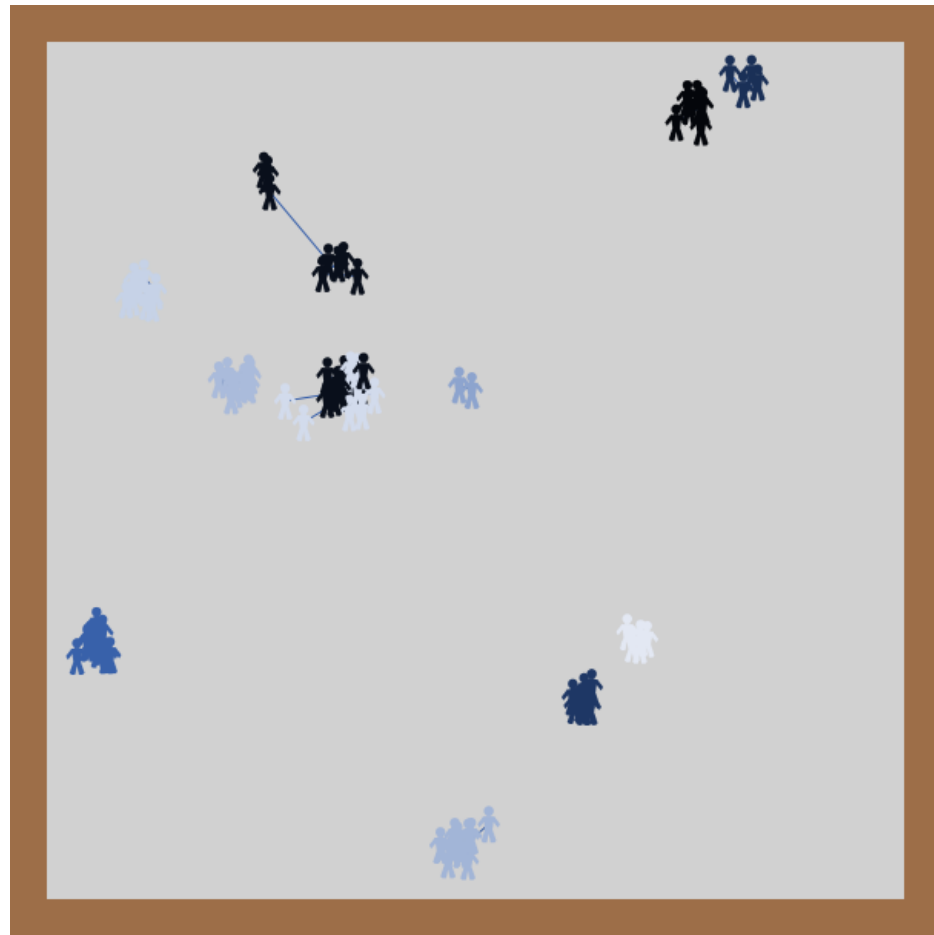


Figure 3.9. Simulation at iteration 3588 (agents shaded by heading)

A Ripley's (1977) K test of the spatial interpoint distances between agents (a statistic for detecting deviations from spatial homogeneity) was implemented to determine if agents were randomly distributed across the grid. Figure 3.10 shows the results for Ripley's K test on agents at the start of the simulation (iteration zero). The observed spatial pattern of our data, as depicted by the blue line, is situated between the 95 % lower (red line) and higher (green line) confidence intervals. This statistically indicates that the interpoint distances contain no variation that we would not expect by chance—the agents are randomly distributed across the spatial grid at the creation of the simulation.

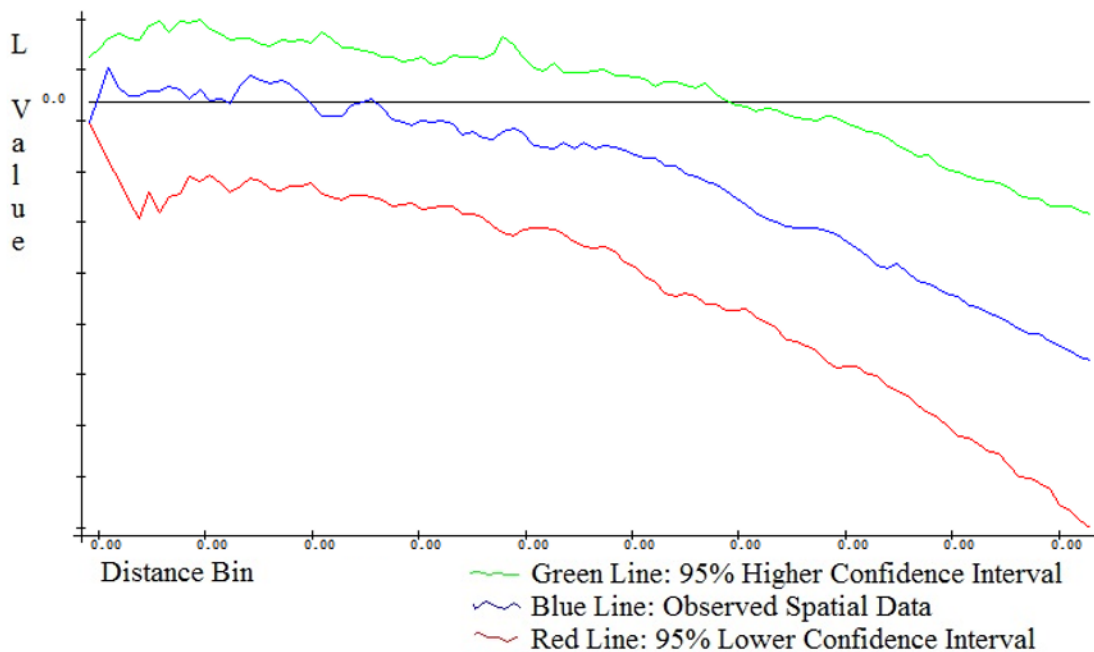


Figure 3.10. Ripley's K at the start of the simulation (iteration zero)

Figure 3.11 shows the results for the Ripley's K test with 1000 simulation runs on our agents after 3588 iterations. The observed spatial pattern of our data, as depicted by the blue line, is situated greatly above the 95 % higher (green line) confidence interval. This indicates a statistically significant clustering of agents at smaller distances than expected by chance ($p < .05$). These results show that sets of agents are now clustered together.

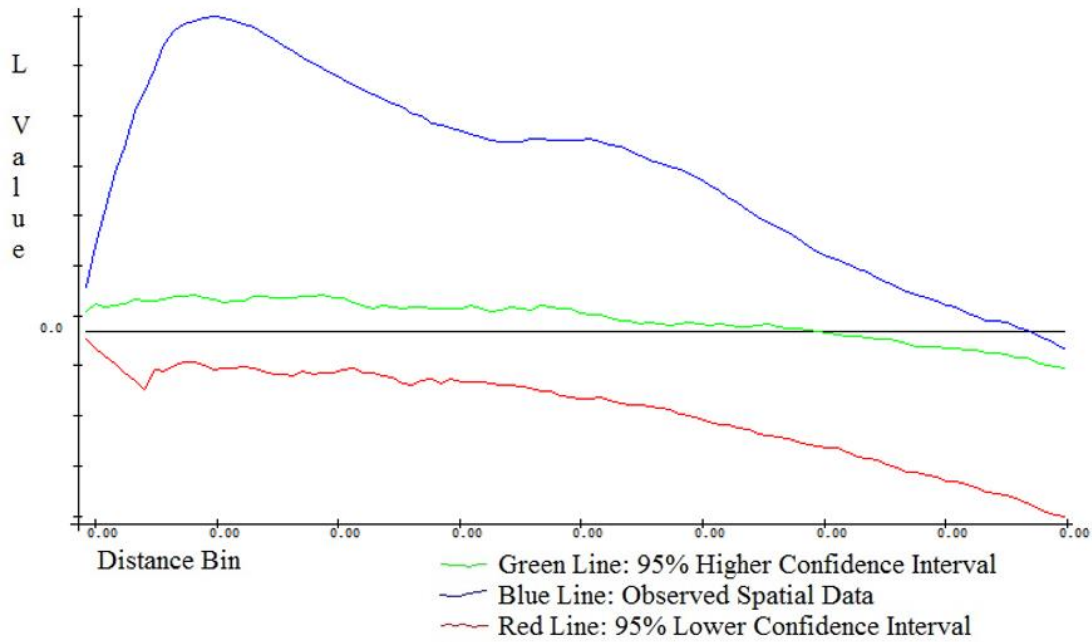


Figure 3.11. Ripley's K test after 3588 iterations

Evidence that the agents are flocking is of interest, but it does not inform us of whether those flocking together are closer in opinion space than we would expect by chance. To address this issue, we conducted a Moran's (1950) I spatial correlation that calculates spatial autocorrelation between location (X-Y coordinates) and a feature value. Essentially, the test determines whether those agents located closer together in physical space (X-Y coordinates) are statistically more likely to have similar characteristics (red and green values) than those at a further distance. Moran's I provides a value from minus one (perfect negative spatial autocorrelation—dispersion) to plus one (perfect positive spatial correlation—clustering), with a p value indicating statistical significance. As with the Ripley's K test, we conducted the first test on the agents at the start of the simulation (iteration zero).

A Moran's I spatial correlational test of agents' red values and locational coordinates at the start of the simulation was non-significant (Moran's $I = -.03$, $SD = 0.02$, $p = .30$, $n.s$). A Moran's I spatial correlational test of the agents' green values and their locational coordinates at iteration zero was also non-significant (Moran's $I = -.01$, $SD = 0.2$, $p = .95$, $n.s$). These

results indicate that at the beginning of the simulation, agents' colour values were neither spatially clustered nor dispersed, but rather distributed randomly across the grid.

Two Moran's I tests for the same agents at iteration 3588 produced significant positive spatial correlations for both red and green dimensions (Moran's I = .53, $SD = 0.04$, $p < .001$ and Moran's I = .55, $SD = 0.04$, $p < .001$ respectively). These results show that after 3588 iterations, low and high values of redness and greenness were more geographically spatially clustered than would be expected by chance.

Although Moran's I is able to distinguish whether those closer together in physical space have similar colour opinions, it ignores the groups defined. More specifically, Moran's I is insensitive as to whether there is sufficient consistency within our numbered physical flocking groups on their colour dimensions, and whether those opinions within a group are equivalent in terms of their absolute value. Therefore, intraclass correlation coefficients (ICC[1] and ICC[2]; Shrout & Fleiss, 1979) were assessed to inspect the relative consistency of colour values within and differences between groups. ICC(1) demonstrates the proportion of variance in colour that can be explained by group membership. ICC(2) establishes how reliably the average colour score for each group can distinguish that group from others.

As there were no groups formed at the creation of the simulation (iteration zero), we did not perform ICCs on this dataset. ICC(1) and ICC(2) for a combined vector of red and green opinions at iteration 3588 found that a significantly large proportion of the variance in colour scores and average colour differences between groups could be explained by group membership (ICC(1) = .94, 95% CI [0.88, 0.98]; ICC(2) = .99, 95% CI [0.98, 1.0], $F(12, 87) = 116.31$, $p < .001$). These intraclass correlation values indicate excellent relative consistency of colour values within groups and large colour differences between groups (Katherine J. Klein et al., 2000). In sum, individuals sharing group membership within our simulations are statistically more similar in physical space (H1a) and opinion space (H1b) than expected by chance.

The next investigation of interest was whether agents migrated between groups, and, if so, whether these migrations could be explained by relative prototypicality of the agent to the local context. To observe migration and each agent's colour preference simultaneously, each agent's movement and colour dimensions were plotted in a video format. In this movie (available at <https://youtu.be/KUszyxZmkEU>), each agent is represented by a coloured bar. The angle of the bar shows the agent's red or green opinion, where horizontal means entirely red and no green, and vertical means entirely green and no red. The length of the bar indicates the strength of the opinion. The bars are also coloured to approximately show the direction of the agent's opinion. Black lines connect agents to their prototypes.

In confirmation of the Ripley's K test of spatial randomness, visual inspection of this video noticeably showed that the agents clustered together in groups. Further, in support of the Moran I's spatial correlation test and the ICC tests, agents appeared to get together in groups with those of similar opinions (as represented by the angle and length of the bars in the video). Throughout the simulation, membership of the groups did not appear to fully settle at any time, but rather agents were seen leaving groups and joining others.¹² It did not appear that large groups schismed purely by being overly large. Rather, it appeared that single agents were 'picked off' by other groups (for example, see most northern red agent at iteration 1503 of uploaded video); or that two groups merged to make one larger group in the presence of a third group, only to then separate again after this third group departed. For example, in Figure 3.12 (iteration 3696), there exists one group in the centre of box A, who are joined together in the presence of a comparatively dissimilar red group (shown in box B). At iteration 3707, this A group remain together in the presence of another comparatively dissimilar green (box C) group. However, once isolated away from these comparatively

¹² Although not included in the thesis (owing to brevity), we compared this 'metacontrast' model to a model in which agents linked not with a context-dependent prototype, but with the agent most similar in colour. The groups in this 'similarity' model stabilised in less than 100 iterations and no migration occurred thereafter. This suggested metacontrast may be an important driver of group migration.

different ‘others’, and as the local context changes, this A group then separate into two distinct groups who move apart.

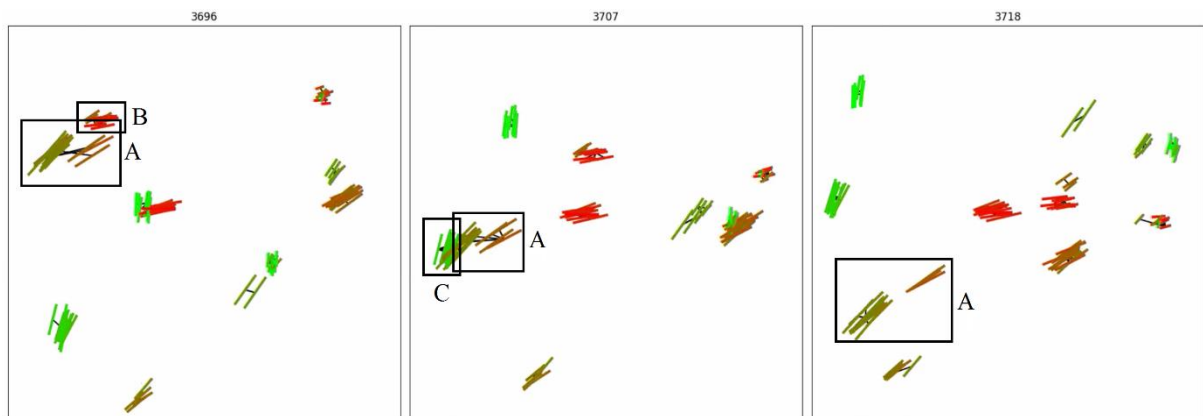


Figure 3.12 Modelling example of group merging and separation

Therefore, in line with SCT predictions, individuals are found to migrate between social groups (H2) through fluctuations in prototypicality and social context.

Two paired-sample *t* tests were conducted to ascertain whether agents who had migrated from a social group to join another were further from their average ingroup members in physical distance and psychological colour distance, and further from their colour prototype, than similarly matched agents who had remained. To this end, we randomly¹³ selected a sample of 250 agents over 5000 iterations, who had migrated (hereafter ‘migrators’) to a new social group. This sample was then compared to an equal-sized sample of 250 ‘remainder’ agents. Each remainder agent was a matched control selected on the basis that it had belonged to the same social group as the ‘migrator’ and was the closest agent in physical distance, but that it had remained in the social group while the matched migratory agent had left. Psychological distance to an agent’s prototype was calculated as the squared difference between each agent’s colour dimension and the colour dimension of its prototype before the ‘migrator’ departed. The psychological distance to an average member of the agent’s group was calculated as the squared difference between each agent’s colour

¹³ Once data from a random ‘migratory’ agent was taken, another ‘migratory’ agent could not be randomly selected for five iterations, thus ensuring independence between those within the sample.

dimension and the mean colour dimensions of those within the social group before the ‘migrator’ departed. Physical distance was calculated as the squared difference between the agent’s spatial coordinates and the median coordinates of social group members before the ‘migrator’ departed.

There was a significant difference between migrators and remainers in the psychological distance between themselves and their prototypes. Migrators were psychologically further in distance from their prototype ($M = 0.06$, $SD = 0.04$, $SE = 0.002$) than remainers ($M = 0.02$, $SD = 0.02$, $SE = 0.002$, $t(500) = 15.46$, 95% CI [0.03, 0.04], $p < .001$, Cohen’s $d = 0.98$). There was also a significantly greater psychological distance between a migratory agent’s colour value and their average groups’ colour value in comparison to the matched remainder agent ($M = 0.03$, $SD = 0.02$ and $M = 0.01$, $SD = 0.01$ respectively, $t(500) = 12.24$, 95% CI [0.01, 0.02], $p < .001$, Cohen’s $d = 0.76$). Finally, although remainder agents were matched as closely as possible in regard to physical distance—i.e., remainder agents were specifically chosen for being the closest of the remaining agents in distance to the migratory agents—migrator agents were significantly further from the physical centre of mass of a social group than remainder agents ($M = 1.89$, $SD = 2.35$ and $M = 0.59$, $SD = 1.65$ respectively, $t(500) = 9.36$, 95% CI [1.02, 1.57], $p < .001$, Cohen’s $d = 0.59$). All of these effect sizes were medium to large.

A forced binominal logistic regression was performed to ascertain whether reduction in group size, psychological distance¹⁴ or physical distance to one’s social group were most predictive of when an agent leaves their social group for another. The binominal logistic regression model was statistically significant ($\chi^2(3) = 194.87$, $p < .001$) and explained 43 %

¹⁴ The original objective was to include both psychological distance to prototype and psychological distance to an average member in the logistic regression model. However, as prototypes are usually representative of the majority of agents within that group, the psychological distance to an average group member and psychological distance to prototype were highly correlated ($r(500) = .78$, $p < .001$) and failed tests of multicollinearity. We thus excluded ‘psychological distance to average member’ as a predictor.

(Nagelkerke $R^2 = .43$) of the variance in agent migratory movement. The model correctly classified 80% of migratory cases.

Psychological distance from prototype was by far the strongest predictor of whether an agent migrated from one group to another (see Table 3.1). As such, those who had a greater opinion distance from their ideal prototype were most likely to select a new prototypical agent from another social group (H3). Increased physical distance from the centre of the group (i.e., being on the physical fringes) was not associated with a greater likelihood of leaving the group for a different one. There was a negative relationship between migration and new group size (i.e., a tendency to leave a group and join a smaller group), although this relationship was marginally non-significant.

Table 3.1

Results from a binomial logistic regression predicting migration from a social group

Model	χ^2 Model	R^2	B	SE_B	Wald	β	p
Migration Yes/No	***	0.43					
Intercept			-1.22	0.28	18.73	0.30	< .001
Psychological distance			50.80	5.22	94.79	< 0.001	< .001
Group size change			-0.06	0.03	3.54	0.94	.060
X-Y centre of distance			0.10	0.06	2.32	1.10	.127

Note: *** $p < .001$; df (3, 497); Accuracy of predictions for model remainers and migrators = 83.6 % and 76.4 %, respectively, B = estimated unstandardized beta

3.4.7 Discussion

In this study, we applied the principle of metacontrast to social group formation and flocking. Agents derived prototypical opinions based on their local context and shared a heading with the agent most prototypical of this position and moved forward. These simple rules clustered agents together in physical space, who then moved together in groups. These agents were not randomly clustered, but those more similar in regard to colour dimension were closer in physical space than those dispersed further away. Agents were far more likely to have closer colour opinions to ingroup members than to members from other social groups. As predicted by SCT, the category boundaries for agents were fluid and prototypes shifted as

the context and relevant social comparison altered (e.g., Hogg, 2004; Hogg & Terry, 2000). This permitted freedom of movement between groups when agents were unsatisfied with their current prototype (Hogg, 2012), and also allowed for two groups to combine and splinter again depending on the entrance of a third group (Oakes, 1996). In this model, being psychologically distant (i.e., unsatisfied by the local prototype of your group), rather than being physically distant (i.e., further in physical space from the centre of your group), predicted migration. Though agents tended to migrate to smaller groups, as predicted by optimal distinctiveness theory—the principle that individuals seek membership of groups that are moderately sized to best minimise isolation while best maximising differentiation (Brewer, 1991, 1993)—this tendency was statistically non-predictive of migration.

It is of course unsurprising that we found metacontrast-predicted results from a metacontrast model. However, this model may be the first to test these principles dynamically over multiple micro-interactions, with fluctuating local neighbourhood compositions and comparative contexts. Further, this simulation represents the first successful effort of creating dynamic groups who flock together in physical space via metacontrast. As such, this model may be useful for other simulation researchers who would like to create context-dependent social groups, as opposed to pre-assigned fixed social groups. For interested researchers, it is important to remember that the number of comparative dimensions can be expanded beyond two if required (we used two for simplicity). It is also possible to adjust the width and aversion parameters depending on one's own specifications. For example, a researcher wanting larger social groups to form may wish to increase the width parameter—the parameter that determines how different an opinion needs to be before it is considered outgroup—to a larger number (Salzarulo, 2004, 2006). Alternatively, a researcher wanting smaller groups may want to decrease the width parameter. Finally, if a researcher would like greater polarisation away from outgroup opinions, then an increase in the aversion parameter would lead to more extreme prototypical opinions.

Although the aim of this model was to create emergent social groups who could flock together in physical space, interestingly, the model produced unexpected emergent behaviour of interest relevant to other threads of research. By moving closer to agents perceived as contextually ‘similar’ and forming groups away from those categorised with ‘different’ prototypes, our recorded behaviours were in line with the behavioural evidence of Novelli, Drury and Reicher (2010), in which participants were found to move closer to other participants perceived as being ingroup, and further from those perceived as outgroup. Further, our findings echo the phenomena detailed in work on self-segregation (Buttny, 1999; Schrieff, Tredoux, Dixon, & Finchilescu, 2005) and resegregation (Clack, Dixon, & Tredoux, 2005; Dixon & Durrheim, 2003) in which micro adjustments in personal spaces towards those perceived as ‘similar’ in spaces—such as in cafeterias, on university campuses, on beaches and in lecture theatres—may result in segregation on a macro level. This suggests that our metacontrast model of flocking may have fruitful applications in other research areas.

An important observation is that all ICC values of this study were extremely high (> .94). These high ICCs exemplify the fact that the dimensions used to assess within-group consistency were the very same colour dimensions from which agents determined group categories. An important reflection, therefore, may be: what kinds of real-life group formations could this model be compared to? We would suggest that these ICCs would be comparable to real-life situations in which group members are asked about the membership of a group whereby the inclusion/exclusion boundaries are clearly defined. For example, ‘Are you a group member of your specific research group? Of the football team you support? Of your nationality or religious group? Of your organisational group in a traditionally structured company?’ Note in these examples that participants are not asked about a distal outcome variable, rather they are asked about the very definition of their group.

Critically, our model was limited in that it only modelled the principle of metacontrast. As such, the model does not take into consideration that some comparison dimensions may be

more socially prestigious, valued or desired than others (McGarty et al., 1992). Further, our model did not give agents any sense of belonging, attachment or commitment to the group (Ellemers, Spears, & Doosje, 2002), thus making migration easy. Finally, we did not furnish our agents with any motivations to identify with others to reduce feelings of uncertainty (Hogg, 2012) or to seek membership of groups that best minimised isolation while best maximising differentiation (Brewer, 1991, 1993). Although the inclusion of these theoretical parameters would allow novel insights into group formation/retention, they would necessitate greater effort and added complexity that falls beyond the scope of this thesis—specifically to computationally model aggression and violence (which will now be discussed).

4 Dyadic and Social Agent-Based Models of Violence: The Roles of Dynamic Collective Inhibition and Polarisation in Group Violence

Established models of aggression—such as the GAM and the I³ theory (Anderson, 1997; Anderson & Bushman, 2002; Finkel, 2014; Slotter & Finkel, 2011)—have identified an impressive list of personal and situational risk factors relevant in determining the likelihood and strength of aggressive impulses. However, and in common with most models of aggression, these models are less effective in describing or predicting when aggressive individuals become violent. The key criticism of these approaches is the focus on the individual's capacity to inhibit their own aggression at the expense of the wider social context. A more dynamic model would allow for the interaction between all relevant agents in the emerging aggressive episode (Levine et al., 2011) and permit dynamic collective inhibition (Levine et al., 2012; Reicher, 1987; Stott et al., 2007, 2001).

Therefore, in this chapter, we test the principles of the I³ theory in an agent-based modelling environment. We develop an alternative approach that amalgamates the principles from these dyadic accounts with recent developments in group dynamics and the social identity tradition (Hogg & Reid, 2006; Salzarulo, 2006; J. R. Smith & Hogg, 2008). We use our modelling environment to compare established models of aggression and violence with our social identity-based model. In Study 4, we dynamically explore how group processes facilitate the escalation from aggression to violence and inhibit aggressive individuals from becoming violent. In Study 5a, we investigate whether a self-categorisation-informed model of violence produces the intergroup hostility biases predicted by self-categorisation theory (e.g., Otten, 2009; Viki & Abrams, 2013) and as demonstrated in our first CCTV behavioural analysis. We provide a final evaluation of our ABM and discuss future modifications. To begin this chapter, we first provide a description of the ABM we will be testing.

4.1 The Model: Purpose and Brief Description

The model description follows the standard ODD protocol for describing individual models and ABMs (Grimm et al., 2006, 2010). The model was implemented in NetLogo v. 5.2.0 (Wilensky, 1999).

4.1.1 Purpose

The purpose of this ABM is to compare dyadic accounts of aggression with an alternative social model in which individuals take into account collective group norms and the dynamic local social context. While this model will consider the properties and interactions of the individuals, it will also highlight the significance of the dynamic local context of violence.

4.1.2 Brief Description

The social model of violence has three layers (for further clarification, see Figure 4.1 and ‘Process Overview and Scheduling’).

Physical Space. At the bottom layer, agents form social groups with whom they move in physical space (as in Study 3). These social group members have greater influence than outgroup members in determining the agent’s normative position towards violence (determined in social space—the middle layer).

Social Space. The middle layer describes changes in normative position towards violence due to interactions with ingroup and outgroup members. The idea is that an agent’s social component (SC; normative position on violence) tends to the prototypical for agents of its group.

Dyadic Space. At the top layer, pairs of agents interact based on our modification of the I^3 theory. The result of each interaction is an aggression value for each of the interacting agents, with this aggression value being converted into a propensity (probability) of being violent (PoV) value. Aggression is calculated as a function of a neighbour’s instigation (provocation), an agent’s disposition (D) and an agent’s prototypical SC. While an agent’s

disposition does not change throughout a simulation, an agent's SC can be affected by group membership and the local normative context.

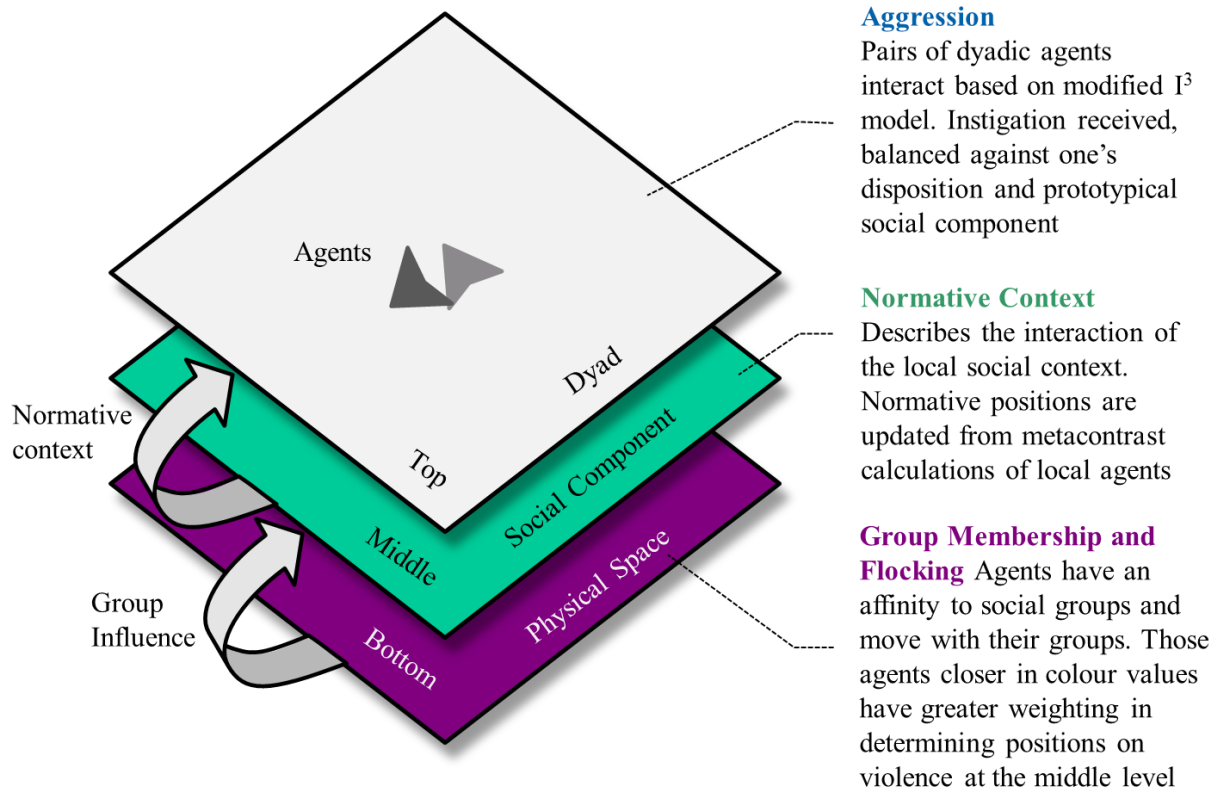


Figure 4.1. Social model of violence

4.1.3 State Variables and Scales

The main entity of this model is the agent. The agent is an individual characterised by two *colour opinions* (as in Study 3), a *disposition* that is an intrinsic property of the agent and a *SC* that depends on factors external to the agent.

The colour opinions are two opinions the agent holds on two separate topics—arbitrarily defined as ‘redness’ (\mathcal{R}) and ‘greenness’ (\mathcal{G}). These opinions range from 0.00 to 1.00 and are randomly assigned from a uniform distribution.

The agent also has an idiosyncratic disposition that is a value randomly assigned from a uniform distribution: 0.00 to 1.00. This disposition is a combination of the impellance and inhibition categories defined in I^3 theory and includes the theoretical factors involved in

aggression, such as trait aggressiveness, aggressive scripts, self-control ability, executive functioning and testosterone. Originally, it was the researchers' intention to differentiate impellance from inhibition as is done in I³ theory. However, when scrutinising the two processes and preparing them for model formalisation, it becomes clear that the categories—and the contents of those categories (i.e., what belongs where)—are somewhat ambiguous (e.g., Galán et al., 2009). This is conceded by the I³ authors who stated:

We raise a caveat: the current state of the literature does not allow for definitive placement of a given risk factor into a particular I³ theory process category. (Finkel & Eckhardt, 2013, p. 460)

We have relied on theory to determine, for example, (a) that social rejection, ego threat, and provocation, are instigating triggers; (b) that testosterone, narcissism, and physical proclivity are impelling factors; and (c) that self-regulatory strength, lack of alcohol consumption, and negative outcome expectancies are inhibiting factors. One limitation of this approach is that existing theory is in many cases not sufficiently developed vis-à-vis the I³ theory parameters to allow for definitive classification. (Slotter & Finkel, 2011, pp. 16–17)

This classification is further hindered by the authors' insistence that each process is orthogonal: “instigation is *distinct* from impellance, impellance is also *distinct* from inhibition” (emphasis added) (Finkel & Eckhardt, 2013, p. 459).

We would argue that while the process of instigation (being provoked) may indeed exist independently, the distinction between impellance and inhibition is far from clear. For example, in their current I³ model, the authors considered low levels of alcohol part of inhibition. However, equally, high levels of alcohol can be considered an impellance and something that facilitates aggression. In a similar fashion, aggressive scripts may be part of

one's impellance and provide a framework of how to behave aggressively. Equally, aggressive scripts can exist as something that reduces one's capability or desire to inhibit their aggressive behaviour. As such, the presence of a factor on one side of the equation may represent the absence of a factor on the other side—and vice versa. This point is exemplified in the following extract from the I³ authors:

We are reasonably confident that dispositional self-control predicts reduced aggression in large part by raising the inhibition threshold ... thereby increasing the likelihood that individuals will override aggressive impulses. That said, perhaps such dispositional self-control also predicts reduced aggression in part by reducing impelling forces, thereby decreasing the strength of the aggressive impulse in the first place. (Slotter & Finkel, 2011, p. 17)

As we cannot identify the difference between impellance and inhibition, we combine both factors into a single disposition to nonviolence. Low dispositions (as opposed to high dispositions) mean that one is more prepared to react aggressively when a provocation (instigation) arises. We consider this disposition stable within a short period and therefore keep the value constant throughout the simulation.

In addition to colour opinion and disposition, each agent has an SC that is a value randomly assigned from a uniform distribution: 0.00 to 1.00. This SC is the agent's perceived normative position on violence and can be adjusted via metacontrast calculations to be prototypical of the agents under consideration.

The sum of the disposition and the agent's SC represents the agent's *response to provocation*. If the sum is small, the agent tends to react to small instigations with a high level of aggression. If the sum is large, the agent needs a large instigation to react aggressively. It is important to note that while an agent's disposition is fixed throughout the simulation, the

agent's SC (i.e., perceived normative position on violence) is fluid and changes in relation to the shifting local social context.

Akin to Study 3, the simulation runs on a square lattice of 25 x 25 cells with a one-cell-thick boundary. As agents are mobile, each agent has an X-coordinate and a Y-coordinate that denotes its location on the physical grid. Each agent also has a heading that determines the direction it will take with the next physical step, an identity number to differentiate it from another, and a local neighbourhood (\mathcal{N}_i) that is the local vicinity in which that agent interacts with others. Similar to Study 3, this neighbourhood is restricted to three ($\mathcal{N}_i = 3$). The model runs in iterations or 'ticks' in the NetLogo language.

4.1.4 Process Overview and Scheduling

The model has three main processes: a) determining social groups and movement (physical space—bottom layer); b) determining the normative position on violence (prototypical SC) (social space—middle layer); and c) determining one's level of aggression and PoV (dyadic space—top layer).

4.1.5 Physical Space: Determining Social Groups and Movement

In order to explore the effect of groups we need to assign agents to social groups. We implement our group flocking metacontrast model from Study 3 to allow the agents to form their own social groups derived from the local 'colour' context. We employ the same methodology as in Study 3 and instruct agents to derive a theoretical colour prototype (the local maximum $P(a, \mathcal{N}_i)$) to link with agent k , the agent most similar in colour opinion to that prototype, and to share a physical heading with that individual. Rather than salient categories, these physical groups represent social groups, or 'colour' opinion groups, who form through contextual perceived similarities on topics, repeatedly interact (for common bond groups, see Prentice et al., 1994; for literature review, see Ren, Kraut, & Kiesler, 2007) and move together in public spaces. These social ingroups have greater influence on an agent's normative

behaviour when an aggressive episode arises. The flocking procedure and all mathematical formulae remained unchanged from Study 3.

4.1.6 Social Space: Determining the Normative Context

The social space (middle layer) is responsible for determining each agent's perceived normative position on violence. An individual's SC value is compared to the SC values of all other agents in the local neighbourhood ($v_{j \in \mathcal{N}_i}$) including outgroup members, via a second distinct metacontrast calculation $P(v, \mathcal{N}_i)$. This determines the prototypical position on violence for that agent. This prototypical SC value (v_i^*) is then combined with the agent's disposition to calculate the agent's level of aggression in dyadic space (the top layer).

We define a function of how prototypical an agent's SC is of the neighbours of i that is given by how large its inter-category distance is compared with its intra-category distance.

This SC prototypicality function is defined as the weighted combination of the two distances:

$$P(v, \mathcal{N}_i) = \alpha \times d_{inter}(v, \mathcal{N}_i) - (1 - \alpha) \times d_{intra}(v, \mathcal{N}_i)$$

where $\alpha \in [0,1]$ is a parameter signifying the outgroup aversion (0 equals no repulsive effect due to outgroup).

We find the mean distance between agent i 's SC and the SC of the ingroup neighbours using the d_{intra} function:

$$d_{intra}(v, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} m(a_j, a_i^*) \mu(v_j, v) \|v - v_j\|^2}{\sum_{j \in \mathcal{N}_i} m(a_j, a_i^*) \mu(v_j, v)}$$

where $\mu(v_j, v) = \exp(-\frac{\|v_i - v\|^2}{\varphi^2})$ is a function measuring the similarity of an agent's SC to the group's prototypical SC. In accordance with previous studies, $\varphi \in [0,1]$ is a width parameter determining the tolerance of differing values (a φ value close to zero signifies that very few different values are considered ingroup values; a value of one signifies that all opinions, regardless of their diversity, are considered ingroup). Although the principle for the normative

context width (φ) is identical to the colour physical space width (ω) (see Study 3), the two parameters are distinct.

Likewise, we can determine the mean distance to the outgroup neighbours' SC using the d_{inter} function:

$$d_{inter}(v, \mathcal{N}_i) = \frac{\sum_{j \in \mathcal{N}_i} m(a_j, a_i^*) (1 - \mu(v_j, v)) \|v - v_j\|^2}{\sum_{j \in \mathcal{N}_i} m(a_j, a_i^*) (1 - \mu(v_j, v))}$$

Similar to previous metacontrast formulae, distance in SC is moderated by a 'fuzzy membership' $\mu(v_j, v)$ function that determines how similar two positions are to one another. However, in the formulae above, the distance in SC is now also moderated by the flocking social group membership to which i belongs. Therefore, while agents take into account the position on violence of all agents within the vicinity when deriving a prototype, outgroup positions have a smaller weighting. Put simply, individuals pay less attention to positions of violence from outgroup members. We denoted the formulae this way because of empirical work suggesting that individuals may be greater influenced and more receptive (though not unconditionally influenced or receptive) to normative information projected from an ingroup source (Turner, 1991; Turner & Oakes, 1986). Furthermore, we did not agree with traditional approaches that advocate a lack of individual agency amongst group members (e.g., Le Bon, 1897/1960; for recent critique on social 'agency', see Swann & Jetten, 2017). Finally, individuals have qualitatively reported that they frequent public spaces on dimensions not related to violence (in our case, colour) and often disagree with fellow group members about the acceptability of violence (Levine et al., 2012). Therefore, in our formalisation, if agents in the neighbourhood of i are close in social group 'colour' opinion space to i , then their SC positions count strongly in the SC distance calculation. Closeness in flocking social group opinion space is measured by $m(a_j, a_i^*)$, which—similar to Study 3—is the 'fuzzy membership' function that determines how similar the other agent is in opinion space to agent i 's prototype.

Once again, the prototypical opinion v^* for agent i is found by locating the local maximum of $P(v, \mathcal{N}_i)$. This is achieved with the same minute adjustments and recalculations as Study 3.

Having found the prototypical position on violence for agent i ,

$$v_i^* = \underset{v}{\operatorname{argmax}} P(v, \mathcal{N}_i)$$

this prototypical SC (position on violence) (v_i^*) is used to calculate the agent's aggression value (as expressed in dyadic space—the top layer below). Therefore, the agent uses its theoretical SC (prototype's position on violence value) rather than its own starting SC. The prototypical SC (position on violence) may not be the position of any specific other agent—it is what the agent calculates ('perceives') the prototypical position of agents in the neighbourhood to be.

4.1.7 Dyadic Space: Determining One's Level of Aggression

The dyadic space (the top layer) describes the aggressive interaction of two agents. The idea is based on a modified I³ theory. In line with I³ theory, agents are provoked by others through an instigation. An agent receives this instigation from any other agent chosen at random within an agent's neighbourhood. This instigation value comes from the previous aggression (formalised as the 0.00 to 1.00 PoV value—see below) of the agent with whom it interacts. As such, an interactant who was very aggressive in the last iteration (i.e., had a high PoV value) will provide a strong instigation (provocation) in the current iteration—and vice versa.

An agent then balances this received instigation against its own *response to provocation* that is the sum of the disposition and the agent's prototypical SC (position on violence) (v_i^*). The result of this calculation is an aggression (A) value.

$$A = \frac{\textit{Instigation}}{\textit{Disposition} + \textit{Prototypical SC } (v_i^*)}$$

$$A = \frac{\textit{Instigation}}{\textit{Response to Provocation}}$$

If the sum in the denominator is small, the agent tends to react to small instigations with a high aggression value. If the sum of the denominator is large, the agent tends to require a large instigation to generate a high aggression value.

Using a piecewise linear function, this aggression value is then transformed (operationalised) into a PoV (0 to 1 bounded probability) value.

$$\textit{Propensity of Violence} \begin{cases} 0 & \textit{if } A < 0 \\ \frac{A}{b}(1 - P^0) + P^0 & \textit{if } 0 < A < b \\ 1 & \textit{if } A > b \end{cases}$$

In this function, the b parameter determines the slope of the conversion line and caps the probability of violence at 1 (see Figure 4.2). The P^0 parameter fixes the Y-intercept and ensures that all agents provide at least some provocation—recall that instigation (provocation) from an interactant comes from that interactant's previous PoV value. By feeding an aggression value into this function, an agent has a chance of being violent that increases with higher levels of aggression; no agent is guaranteed to be neither violent nor peaceful, unless they meet and exceed the b parameter threshold. The slope is set to be gradual, which means that the majority of agents are peaceful (as they tend to be in NTE contexts; Hobbs, Hadfield, Lister, & Winlow, 2003; Hobbs et al., 2005; Levine et al., 2012) and violent outbursts are approximately 10 %.¹⁵

¹⁵ 10 % is somewhat arbitrary and an investigator could realistically increase or decrease this figure. However, we assumed that 10 % (while still on the relatively peaceful side) would allow enough violence to sufficiently analyse.

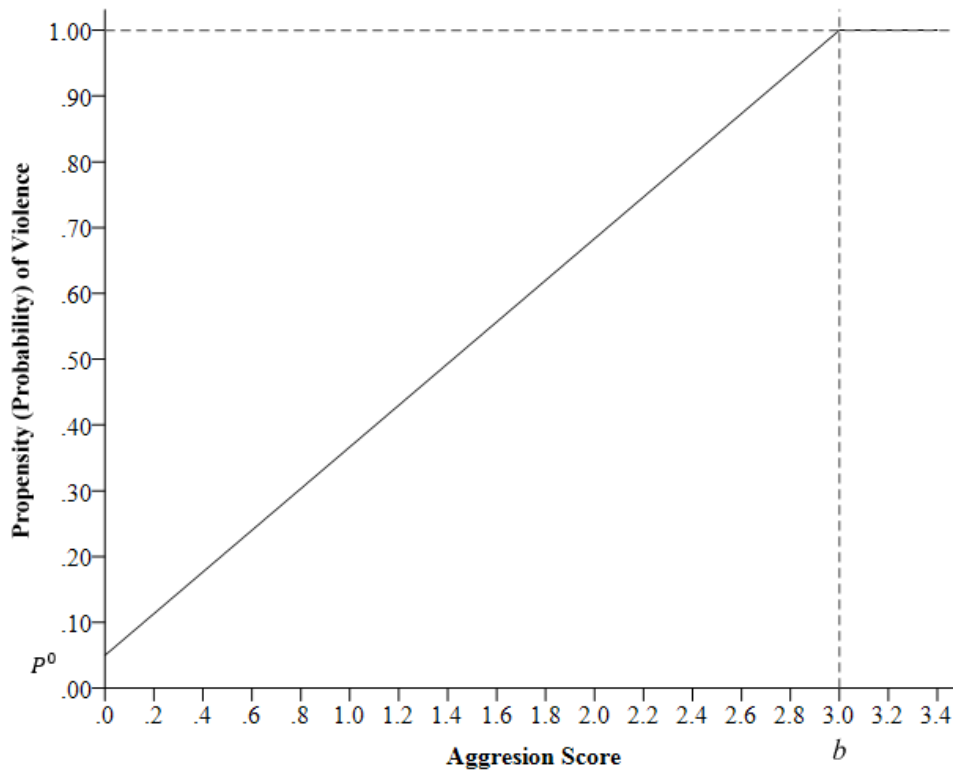


Figure 4.2. Linear function mapping agent aggression to PoV
(Here, $P^0 = .05$, $b = 3.0$; recommended parameters: $P^0 = .05$, $b = [2.5-3.5]$)

4.1.8 Model Steps

Agents are updated once per tick according to the following steps:

- 1) Identify the agents in the neighbourhood \mathcal{N}_i of agent i (including self).
- 2) Form a colour social group and take a shared heading as in Study 3.
- 3) Derive the location v^* from the SC of all agents within \mathcal{N}_i —this is the local maximum $P(v, \mathcal{N}_i)$ that is the closest in SC space to v_i . As such, this is the prototypical position on violence for i .
- 4) Select any other agent at random from within the local neighbourhood \mathcal{N}_i of agent i . Receive an instigation from this individual and calculate an aggression value that is operationalised as a PoV value. Set colour shading to this PoV value.
- 5) Once all agents have completed step 4, take one step forward towards colour social group prototype as described in Chapter 3 and repeat steps 1–5.

4.1.9 Design Concepts

Emergence. Social groups, physical flocking and contextual positions of violence will result from the metacontrast calculations of the single agents. Collective inhibition and violence will emerge from the evolving local context.

Adaptation. An agent will determine a perceived normative position on violence from their local normative context. These positions on violence are not fixed, but rather context dependent and alter with the comparative situation. Agents will adopt this derived normative position on violence for the iteration.

Objectives. The objectives of any agent are to locate and move with a colour social group, determine a perceived normative position on violence based on the local context, provide instigation to a neighbouring agent and calculate an aggression or PoV value from the instigation received by another.

Sensing. Agents are assumed to know the colour opinions, group IDs (to identify if an agent is ingroup or outgroup), SCs (to derive prototypical positions on violence) and PoV values (equivalent to received instigation) of all agents (including themselves) who reside within their neighbourhood. The agent determines which colour opinions and positions of violence are prototypical using metacontrast functions.

Interaction. The agents interact both indirectly and directly. Agents interact indirectly using their own and others' colour opinions and SCs to determine their prototypical positions. Agents also directly provoke one another with an instigation (their previous PoV value) and potentially respond to a received provocation with aggression and violence.

Stochasticity. The order in which agents calculate their metacontrast calculations is pseudo-random. Agents devise their headings and take one step forward simultaneously in a pseudo-random ordering.

Observation. In addition to the same information exported from Study 3 (see Study 3 ‘Observation’), this model also saves each agent’s starting SC, prototypical SC (v_i^*), absolute difference between starting SC and prototypical SC, disposition, aggression value, PoV value and record of the number of agents within the neighbourhood. The disposition of each agent’s interactant and the instigation value received from this interactant are also recorded. Each visual output frame is saved and then combined with others as a simulation movie. The NetLogo world is saved in its entirety and can be returned to for further scrutinisation or data abstraction.

4.1.10 Initialisation, Inputs and Submodels

Simulations are initialised with a population of 100 agents. All agents are created with a uniformly-distributed random redness and greenness opinion between 0.00 and 1.00. Each agent is given a uniformly-distributed disposition value and uniformly-distributed SC value that are both between 0.00 and 1.00. Each agent is placed on the square 25 x 25 lattice with a random X-Y coordinate starting location and a random heading. Similar to Study 3, the neighbourhood for each agent is set at three to ensure interactions only with those agents who are locally present. We set the width (ω , φ) and aversion (γ) parameters for the SC metacontrast calculations at 0.36 and 0.08 respectively, as calibrated by Salzarulo (2004) using the experimental data of Haslam and Turner (Haslam & Turner, 1995) (for discussion, see Salzarulo, 2004). The model does not include any external input data driving environmental factors. The model does not include any submodels.

4.2 Study 4: A Comparison of Social and Dyadic Models of Violence

In this study, we use our modelling environment to compare a social model of violence—which includes updating each agent’s SC to the local normative context—with a dyadic model of aggression in which the local normative context remains constant for each agent (see Figure 4.3). Our goal is to explore dynamically whether group processes can

facilitate escalation from aggression to violence, and also inhibit aggressive individuals from becoming violent as predicted by the social identity tradition.

In the social identity tradition, Postmes and Spears (1998) showed that, rather than amplifying a general norm (which facilitates the expression of violence), shared group identity can lead to the expression of behaviours consistent with situation-specific norms. They argued that these norms are shaped by locally-salient social identities, and that the content of these identities can favour prosocial as well as antisocial actions. In other words, it is the norms of the local identities that shape whether it is acceptable for individual group members to become violent. CCTV behavioural analyses and NTE participant interviews (Levine et al., 2012, Study 1, 2011) provide support for this group-based inhibition account and suggest that group members ‘police [their] own’ to deter aggressive deviants from acting violently. Comparatively, in dyadic accounts of aggression (Anderson, 1997; Anderson & Bushman, 2002; Finkel, 2014; Slotter & Finkel, 2011), the roles of group membership and local social context in inhibiting or facilitating violence are neglected.

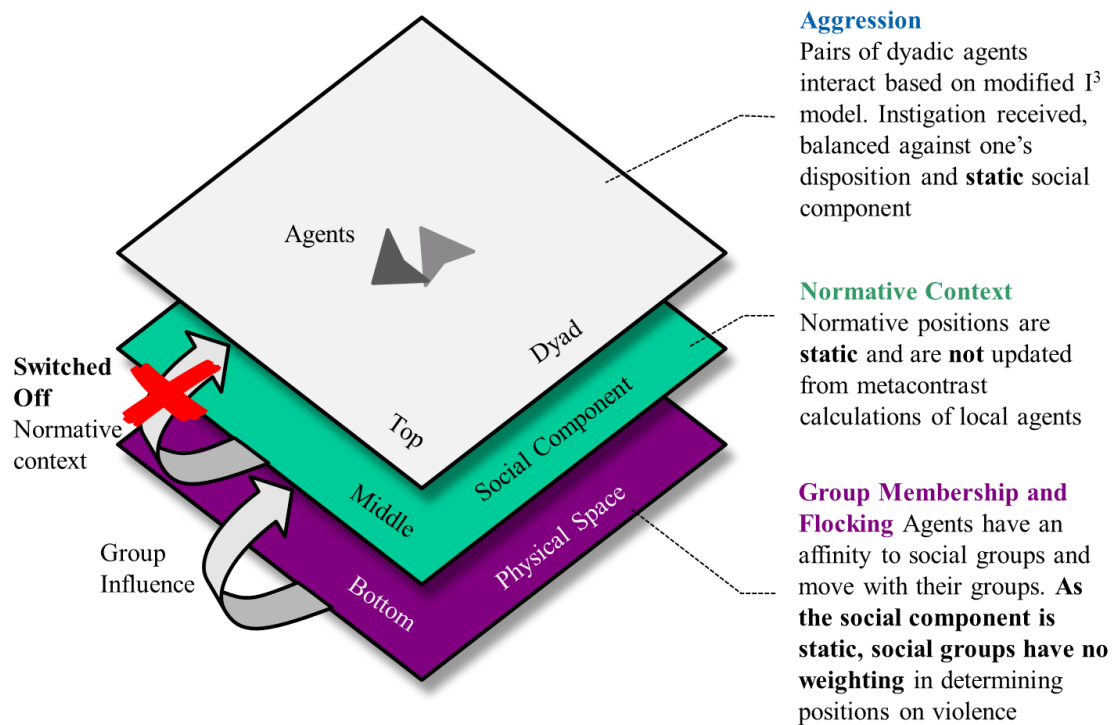


Figure 4.3. Dyadic model of violence (modifications from the social model of violence are in bold)

4.2.1 Hypotheses

H1: We predict that both violence and nonviolence will be spatially more clustered in the social model than in the dyadic model. This is because disposition can be inhibited or uninhibited by the local social context in the social model, whereas in the dynamic model, each individual will act based on their individual characteristics and ignore the local social context.

H2: We predict that the level of violence will be more group normative (i.e., similar within groups) in the social model than in the dyadic model (H2a) and that social groups differ more in their normative levels of violence in the social model than the dyadic model (H2b).

4.2.2 Method

To investigate these predictions, we repeat an identical simulation with one slight modification. In our first simulation—using the social model—we allow our agents at each iteration to determine their perceived normative position on violence from the neighbours in their local context and to adopt this position when calculating an aggression value.

$$\textit{Social Model Aggression} = \frac{\textit{Instigation}}{\textit{Disposition} + \textit{Prototypical SC} (v_i^*)}$$

In our alternative simulation—in line with the dyadic model—we keep the SC of our agents constant and do not update their position at any point in the simulation. As such, the SC of an agent in the dyadic model is static, non-context dependent and acts as a second disposition. This fixed SC is then adopted when calculating an aggression value.

$$\textit{Traditional Model Aggression} = \frac{\textit{Instigation}}{\textit{Disposition} + \textit{Fixed Social Component} (SC)}$$

All agent values (disposition, starting SC, spatial coordinates, and redness and greenness) are identical in both simulations. The order of agent–agent interactions—and as a result, the colour social group formation and movement—are also identical for both models. All model parameters (width values, aversion values and linear function parameters) are identical in both simulations. This allows duplicate simulations with the only variation being whether an agent takes into consideration the normative position on violence of those around them in determining their own position on violence or not.

4.2.3 Results

Figures 4.4 and 4.5 depict the world generated at set-up (iteration zero). Both simulations have identical set-ups and thus appear the same.

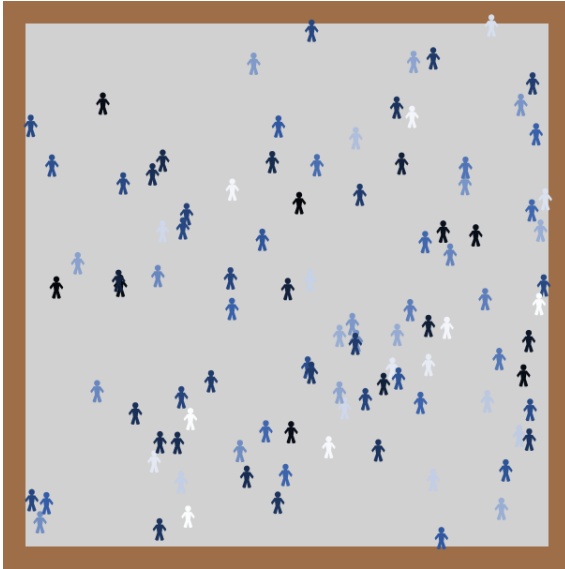


Figure 4.4. Tick zero—Norms constant

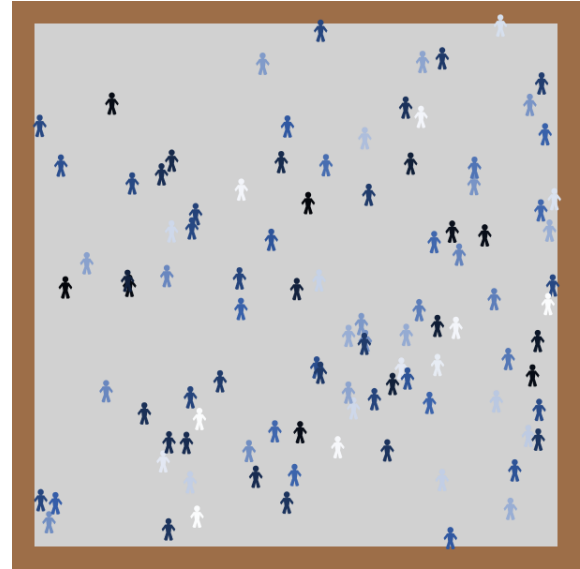


Figure 4.5. Tick zero—Norms update social model. Shaded by heading

Figures 4.6 and 4.7 show the world after 4267 iterations—a random time point between iterations 2500 and 5000 (see below footnote 15 for details on randomisation). Akin to our flocking Study 3, agents appear clustered together in groups that create larger social contexts. Agents are physically distributed identically in both simulations as we would expect. Also, the majority of agents across both simulations are a very light shade of red, indicating they are quite peaceful (as expected given the gradual PoV linear slope). However, visually there appears to be qualitative differences between simulations in regard to the dispersion of agents who have higher PoVs (as denoted by the darker shadings of red).

For example, in caption A1 of the dyadic model, one dark-shaded agent to the right of the box is extremely aggressive. The other surrounding agents tend to be mild in aggression, or mildly/moderately aggressive—thus, denoting local disparities. For the equivalent set of agents in the social model (caption B1), this highly-aggressive agent is considerably more peaceful (as denoted by the comparatively lighter shading), and the neighbour agents within the outlined box have similar PoVs, as indicated by the shadings.

In caption A2 of the dyadic model, there appears to be a dyad consisting of an aggressive agent (depicted in a shade of brown) interacting with another fairly aggressive

agent (depicted in orange). In caption B2 of the social model, the equivalent dyad is considerably more peaceful (lighter shading), with surrounding agents slightly more aggravated (depicting darker shadings than their equivalents in A2). As such, there appears to be more relative consistency of PoVs among neighbours in the social model than the dyadic model.

In caption A3 of the dyadic model, there is one particularly aggressive agent depicted in a shade of dark brown and one other aggressive agent depicted in dark orange. The other agents have a range of varying PoVs. In caption B3 of the social model, the equivalent agents are still the most aggressive (darkest shadings) in the neighbourhood; however, they are a lighter shade than their A3 equivalents. Interestingly, beyond the dyad, the surrounding agents within the neighbourhood (A3) appear to be relatively similar in PoVs. In A3, we can again visually infer that agents within close proximities in the social model are reaching relatively better consistency on their PoVs than their equivalents in the dyadic model.

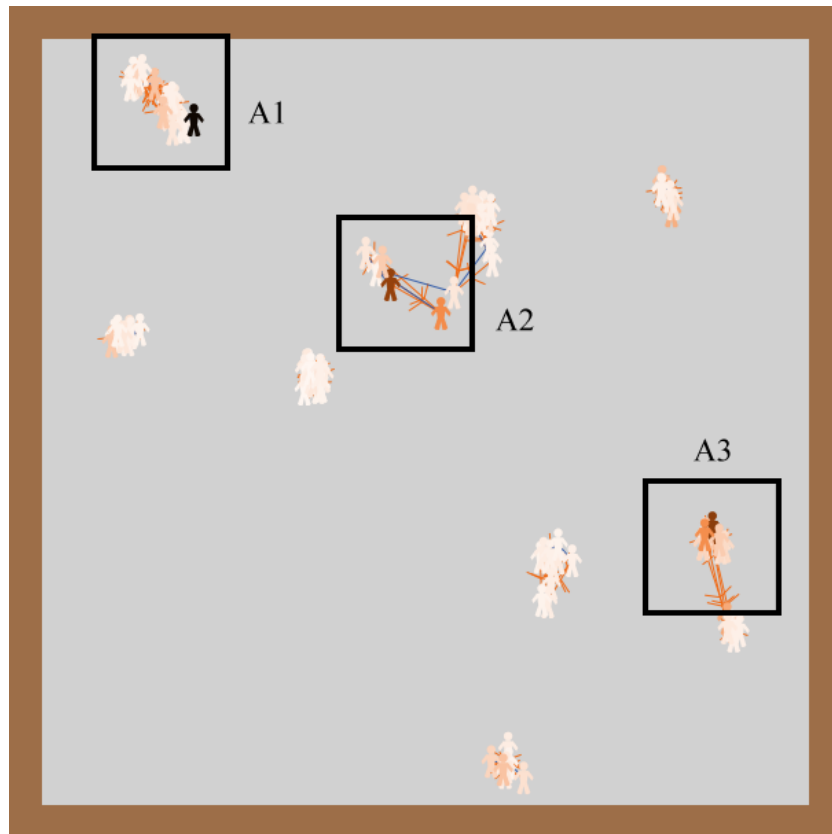


Figure 4.6. Iteration 4267—Norms constant dyadic model

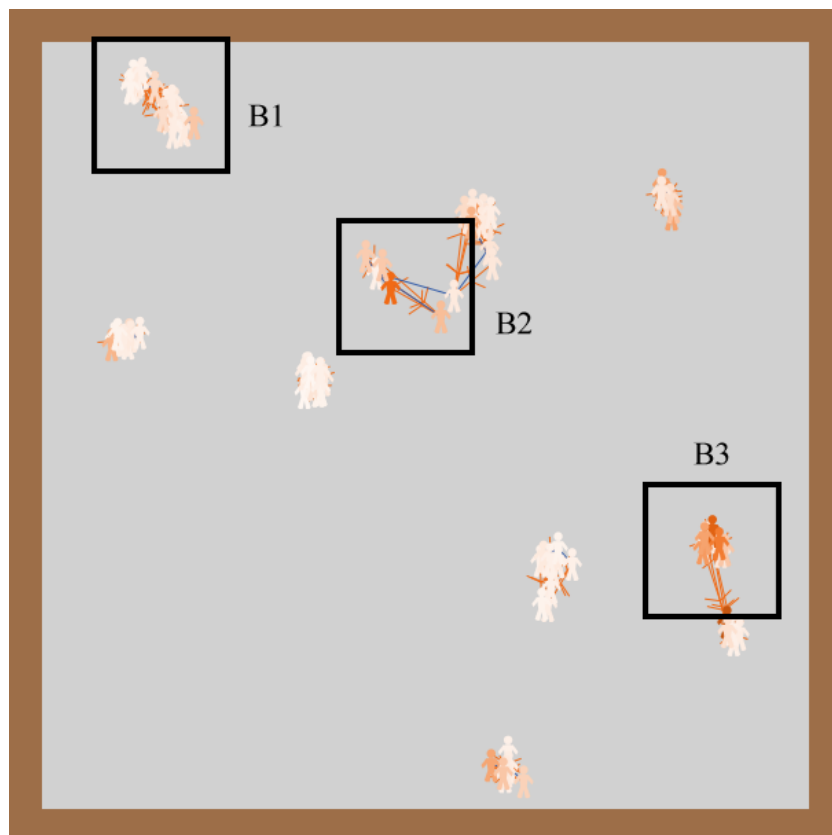


Figure 4.7. Iteration 4267—Norms updated social model.

To test Hypothesis 1 (both violence and nonviolence are more spatially clustered in the social model than the dyadic model), Moran's I 's spatial correlation tests were conducted. This test statistically determined whether those agents located closer together in physical space were more likely to have similar PoVs at 12 time points¹⁶ for both the dyadic and social models.

Table 4.1

Moran's I auto-spatial correlation for X-Y coordinates and PoV for two models

Iteration	Dyadic Model			Social Model		
	Moran's I	<i>SD</i>	<i>p</i>	Moran's I	<i>SD</i>	<i>p</i>
2500	.16***	0.04	< .001	.13***	0.04	< .001
2654	.02	0.03	.39	.12***	0.03	< .001
2984	-.02	0.03	.65	.04	0.04	.19
3075	.05†	0.03	.08	.05*	0.03	.03
3359	.01	0.03	.56	.17***	0.04	< .001
3601	.06*	0.03	.03	.07*	0.03	.02
3695	.04	0.03	.11	.15***	0.02	< .001
4258	.02	0.03	.38	.03†	0.02	.07
4267	.04	0.04	.17	.10**	0.04	.004
4361	.09**	0.03	.004	.18***	0.03	< .001
4744	.02	0.03	.21	.17***	0.04	< .001
5000	.01	0.03	.41	.03*	0.02	.03

Note. *N* of agents for each test = 100; † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

The majority of Moran's I tests for the dyadic model (nine out of 12 sampled time steps) found no statistically significant spatial correlations between physical coordinates and PoVs ($p > .050$; see Table 4.1). As such, at these nine time points, the observed spatial pattern in PoV values were randomly distributed across the spatial domain. At iterations 2500, 3601

¹⁶To ensure that the model had settled before collecting data, we only considered time points after 2500 iterations. From the twelve time points, two of these time points were iterations 2500 and 5000 (the start and end of data recording). Ten of these time points were randomly generated numbers between 2500 and 5000. To get an even spread of numbers, two of these time points were randomly generated numbers between 2500 and 3000, four of these time points were randomly generated numbers between 3000 and 4000, and four of these time points were randomly generated numbers between 4000 and 5000.

and 4361 (three out of 12 iterations), there was evidence of significant positive spatial correlations between locational coordinates and PoVs for the dyadic model. These results show that at these three time points, high and low PoVs were more spatially clustered than would be expected by chance.

Comparably, the majority of Moran's I tests for the social model (10 out of 12 sampled time steps) found significant positive spatial correlations between physical coordinates and PoVs ($p < .050$; see Table 4.1). Therefore, at these 10 time points, high and low PoVs were more spatially clustered than would be expected by chance. At iterations 2984 and 4258 (two out of 12 sampled time steps), there was no evidence of significant positive spatial correlations between locational coordinates and PoVs. These results show that at these two time points, the observed pattern in PoV values were randomly distributed across the spatial domain. Taken together, these results indicate that there was a greater likelihood of spatial clustering of PoV values in the social model compared with the dyadic model. It is important to recall that each agent's geographical location, disposition and sequence of interactions are identical for both models. The only modification is that one model (social model) contains normative updating through metacontrast calculations of the local context and the other (dyadic model) does not.

To test Hypothesis 2 (PoV values are more similar within groups [H2a] and more different between groups [H2b] in the social model than the dyadic model), ICC(1) and ICC(2) (Shrout & Fleiss, 1979) were assessed. ICCs inspect the relative consistency of PoV values within and between groups. ICC(1) is used to test H2a and demonstrates the proportion of variance in PoV values that can be explained by group membership. ICC(2) is used to test H2b and establishes how reliably the average PoV value for each group can distinguish that group from others that are available.

The average proportion of variance explained by group membership (ICC[1]; see Table 4.2) over these 12 tests for the social model was 19.5 % ($SD = 12.6$), indicating a

medium effect size (LeBreton & Senter, 2008), and 6.8 % ($SD = 6.1$) for the dyadic model, indicating a small effect size, which supports Hypothesis 2a. The average ICC(2) value was .59 ($SD = 0.25$) for the social model (see Table 4.3), suggesting a moderate ability to reliably distinguish groups based on the group PoV averages (Portney & Watkins, 2009), and .31 ($SD = 0.22$) for the dyadic model, suggesting poor reliability. In total, eight out of the 12 ICC(2) accompanying ANOVA tests conducted (to assess statistical heterogeneity between groups) at the same time points for the social model were significant, compared to only three out of 12 ICC(2) tests for the dyadic model (see Table 4.3). These results broadly support Hypothesis 2b: social groups differ more in their normative levels of violence in the social model than the dyadic model (H2b).

Table 4.2

Comparison of ICC(1)s depicting homogeneity within groups on PoV for the dyadic and social models

Iteration	Dyadic Model	Social Model	Difference (+/-) between Models
	ICC(1)	ICC(1)	ICC(1) Difference
2500	.16	.31	+.15
2654	.03	.20	+.17
2984	< .001	.00	+.001
3075	.06	.09	+.03
3359	.00	.33	+.33
3601	.12	.20	+.08
3695	.03	.14	+.11
4258	.05	.09	+.04
4267	.09	.22	+.13
4361	.19	.42	+.23
4744	.07	.29	+.22
5000	.02	.05	-.03
Average	.07	.20	+.13

Note. N of agents for each test = 100; mean number of groups over iterations = 12, mean number of agents in group = 8.3 (min. = 3, max. = 16)

Table 4.3

Comparison of ICC(2)s depicting heterogeneity between groups on PoV for the dyadic and social models

Iteration	<i>Dyadic Model</i>			<i>Social Model</i>			Difference (+/-) between Models
	ICC(2)	<i>F</i>	<i>p</i>	ICC(2)	<i>F</i>	<i>p</i>	ICC(2) Difference
2500	.62	2.60**	.007	.79	4.67***	< .001	+.17
2654	.17	1.21	.19	.66	2.96**	.002	+.49
2984	< .001	0.76	.69	< .001	0.90	.55	+.00
3075	.35	1.53	.14	.48	1.94†	.05	+.13
3359	.01	1.01	.44	.80	5.09***	< .001	+.79
3601	.53	2.15*	.02	.68	3.11**	.001	+.15
3695	.19	1.23	.28	.56	2.29**	.01	+.37
4258	.32	1.47	.16	.45	1.84†	.06	+.13
4267	.45	1.81†	.06	.70	3.38***	< .001	+.25
4361	.68	3.16**	.002	.87	7.53***	< .001	+.19
4744	.39	1.64	.11	.79	4.76***	< .001	+.40
5000	.11	1.13	.35	.28	1.38	.19	+.17
Average	.32			.59			+.27

Note. *N* of agents for each test = 100; mean number of groups over iterations = 12; mean number of agents in group = 8.3 (min. = 3, max. = 16); †*p* < .10, **p* < .05, ***p* < .01, ****p* < .001

In Table 4.2, we find that social group members do not always reach similar levels of PoVs. This may potentially be explained further upstream by a disparity between social group members regarding their prototypical positions of violence. To investigate this claim, we conducted a brief secondary data analysis and compared the ICC(1) and ICC(2) for the PoV values against the corresponding prototypical SC (v_i^*) ICCs at the same time point for the social model (see Table 4.4). Visual inspection suggests a large correspondence between the consistency and inconsistency within groups on the PoVs and the consistency and inconsistency within groups on the prototypical positions towards violence (see Table 4.4). Statistically, there is a positive relationship between the PoV ICC(1) values and the prototypical SC ICC(1) values (Pearson's $r = .74$, $N = 12$, $p = .01$). As such, when there is poor consistency within groups on a prototypical position of violence, there also tends to be poor consistency on PoVs.

Table 4.4

Comparison of ICC(1)s depicting homogeneity within groups for PoV and prototypical

Iteration	Social Model: Propensity of Violence	Social Model: Prototypical Social Component	Difference (+/-) between Models
	ICC(1)	ICC(1)	ICC(1) Difference
2500	.31	.19	-.12
2654	.20	.21	+.01
2984	< .001	< .001	+.00
3075	.09	.06	-.03
3359	.33	.20	-.13
3601	.20	.13	.07
3695	.14	.20	+.06
4258	.09	.25	+.16
4267	.22	.23	+.01
4361	.42	.29	-.13
4744	.29	.24	-.05
5000	.05	.05	.00
Average	.20	.17	-.03

Note. N of agents for each test = 100; mean number of groups over iterations = 12;
mean number of agents in group = 8.3 (min. = 3, max. = 16)

4.2.4 Discussion

In the current study, we compared a more ‘dyadic’ model of aggression and violence with an alternative social model in which individuals derived prototypical positions on violence from the dynamic local social context. In line with predictions, we descriptively¹⁷ found more spatial clustering of violence and peace in the social model, as individuals were either inhibited or uninhibited by the local social context. We further found that PoV levels varied more across groups in the social model, compared with the dyadic model, with stronger intragroup consistency within groups.

These findings suggest that an adjustment to the prototypical position of one’s local social context can result in normative PoVs within groups and distinctly different PoVs between social groups. This may result in small concentrated areas of violence and larger areas of collective inhibition restricting violence escalation. These findings are in line with the theoretical predictions of the social identity literature (e.g., Drury & Reicher, 1999; Reicher, 1984) and the behavioural evidence of Levine et al. (2011).

Conversely, for the dyadic model, we found a weak effect of within-group and between-group consistency on PoVs on the majority of occasions. Interestingly, even in this more ‘dyadic’ model, we found some evidence of significant geographical clustering of violence and peace, and group-level consistency on this dimension—although this occurred much less frequently than in the social model. One explanation for this finding is that the dyadic model is not entirely ‘dyadic’ and still contains a social element. Specifically, while the agents in the dyadic model did not determine positions on violence according to those around them, the agents did move around together with other agents in social groups (see Section 4.1.5, ‘Physical Space: Determining Social Groups and Movement’). As interactions were

¹⁷ We recognise these models are at present evaluated descriptively rather than inferentially. While inferential tests across models would undoubtedly strengthen the reliability and robustness of our findings, we currently do not know of any appropriate methods that could aid in this endeavour

restricted to the local vicinity (\mathcal{N}_i) and individuals flocked together, each agent tended to reinteract with the same individuals over time, who were usually fellow group members. This is important because each interaction does not occur in vacuum. Rather, the instigation (provocation) received from another agent is operationalised as that provocateur's previous PoV value carried over from the previous iteration. If the instigation received is high, then an individual agent is likely to respond to this provocation with a higher PoV value—and vice versa if low. This higher/lower value then feeds into the next interaction that tends to further escalate or de-escalate levels of hostility within that local vicinity. This arrival at similar levels of violence from repetitive reinteractions parallels the predictions of the violence escalation cycle (VEC; Anderson & Carnagey, 2004) that assumes violence is the result of escalating retaliation between two parties.

Another notable finding from the main study is that agreement on PoVs within and between groups was not always present in the social model. One explanation for this is that agents formed mobile social groups based on colour dimensions of redness and greenness, rather than normative positions of violence. Although the mathematical formula used to derive a prototypical position on violence contains a colour group weighting (i.e., individuals pay more attention to opinions of violence from ingroup members), this does not guarantee that individuals will reach consensus with all other group members. Rather, when normative position of violence becomes the dimension of comparative context, individuals may share prototypical positions with those around them who are not colour social group members. The model was intentionally coded this way on the basis that a) individuals have qualitatively reported that they frequent public spaces as friendship groups (e.g., on dimensions not related to violence) and often disagree with fellow group members about the acceptability of violence (Levine et al., 2012), and b) because of the recent critique of the traditional approach that assumes group members are devoid of agency (e.g., Le Bon, 1897/1960; for recent critique on social 'agency', see Swann & Jetten, 2017).

If social group members are unable to reach similar prototypical positions on violence, then one might expect this disparity to result in poorer consistency in PoV values further downstream. This is what we found in a secondary exploratory analysis. More specifically, when there was poor consistency within groups on a prototypical position of violence, there also tended to be poor consistency on PoVs. This analysis suggests that non-group consistency on positions of violence can at least still be partly explained by group processes—specifically, recategorisation away from colour social group members whose positions are too disparate. Even with this caveat in mind, a comparison of the direction of effects largely supports our predictions and places the social model as a theoretically interesting alternative to the dyadic model.

4.3 Study 5a: Investigation of Intergroup Hostility in Social Model of Violence

The purpose of this study was to investigate whether a self-categorisation informed model of violence would produce the intergroup hostility biases predicted by self-categorisation theory (e.g., Otten, 2009; Viki & Abrams, 2013) and demonstrated in our first CCTV behavioural analysis. To this end, we performed a secondary data analysis on the outputs generated from the social model in the previous study.

Theorists have predicted that polarisation away from comparatively more law-abiding others pushes petty criminals towards more severe crimes (Viki & Abrams, 2013) and terrorist groups towards more extreme attitudes of violence (Borum & Pynchon, 1999; M. A. Wilson, Bradford, & Lemanski, 2009). We expected that intergroup interactions in our model would be significantly more hostile than intragroup interactions, owing to a greater polarisation of prototypical positions of violence in intergroup interactions pushing more extreme positions on violence. More specifically, when individuals determine a prototypical position on violence during an intragroup interaction, they are most likely to reach a shared consensus that is not too distant from their initial position (see previous social ICC(1)s where individuals largely had significant agreement with members of their own group). Conversely, in

intergroup contexts, individuals and their social groups will likely polarise their positions on violence away from the outgroup comparison, thus resulting in more extreme positions on violence.

4.3.1 Hypotheses

H1: We predict that there will be significantly more polarisation in intergroup interactions than in intragroup interactions.

H2: We also predict that the PoV derived from intergroup interactions will be significantly higher than the PoV derived from intragroup interactions (an intergroup hostility bias).

H3: We finally predict that this intergroup hostility bias will be mediated by the larger polarisation in intergroup interactions.

4.3.2 Method

We analysed a secondary dataset outputted from the social model of the previous study. Similar to Study 4, we were cautious to only include interactions that occurred once the model was deemed to have settled between iterations 2500 and 5000. We were left with a sample of 250,000 interactions (100 agents interacting once per iteration for 2500 iterations). Of the 250,000 interactions, 179,991 (72 %) were intragroup interactions and 70,089 (28 %) were intergroup interactions. From these 250,000 interactions, we randomly selected 1000 intragroup interactions and 1000 intergroup interactions for data analysis.

4.3.3 Results

We first tested Hypothesis 1 to see whether agents engaging in intergroup interactions polarised their SC positions on violence more than agents engaging in intragroup interactions on average. This polarisation was operationalised as the absolute difference between an agents' original position on violence (SC) and the new prototypical position (v_i^*) adopted after the metacontrast calculations of the local context. Initial inspection of the descriptive statistics showed that on average, an agent's prototypical position on violence (v_i^*) (bounded

between 0.00 and 1.00) had shifted around 0.15 (from the original SC) after adopting the prototypical position of the local context ($M = 0.15$, $SD = 0.18$). The large standard deviation of 0.18 suggests wide variance between individual cases.

Figure 4.8 indicates that agents shifted away from their original position on violence in both intergroup and intragroup interactions; however, the variance—as represented in the quartile range—appeared slightly larger for intergroup interactions. We conducted a Fligner-Killeen’s test of homogeneity of variances (Fligner & Killeen, 1976) to evaluate whether this difference in variation was significant—that is, whether there was a statistically greater divergence in polarisation values for intergroup interactions than intragroup interactions. We found that the levels of variance for intergroup and intragroup polarisations were statistically different ($\chi^2(1, N = 2000) = 4.65, p = .031$).

Descriptive statistics showed that on average individuals in intergroup interactions ($M = 0.16$, $SD = 0.13$) shifted further away from their original position on violence after prototype calculations than individuals in intragroup interactions ($M = 0.14$, $SD = 0.12$). A Wilcoxon signed-rank test for related samples found this difference to be statistically significant, with a weak effect size ($Z = -3.07, p = .002, r = -.07$).

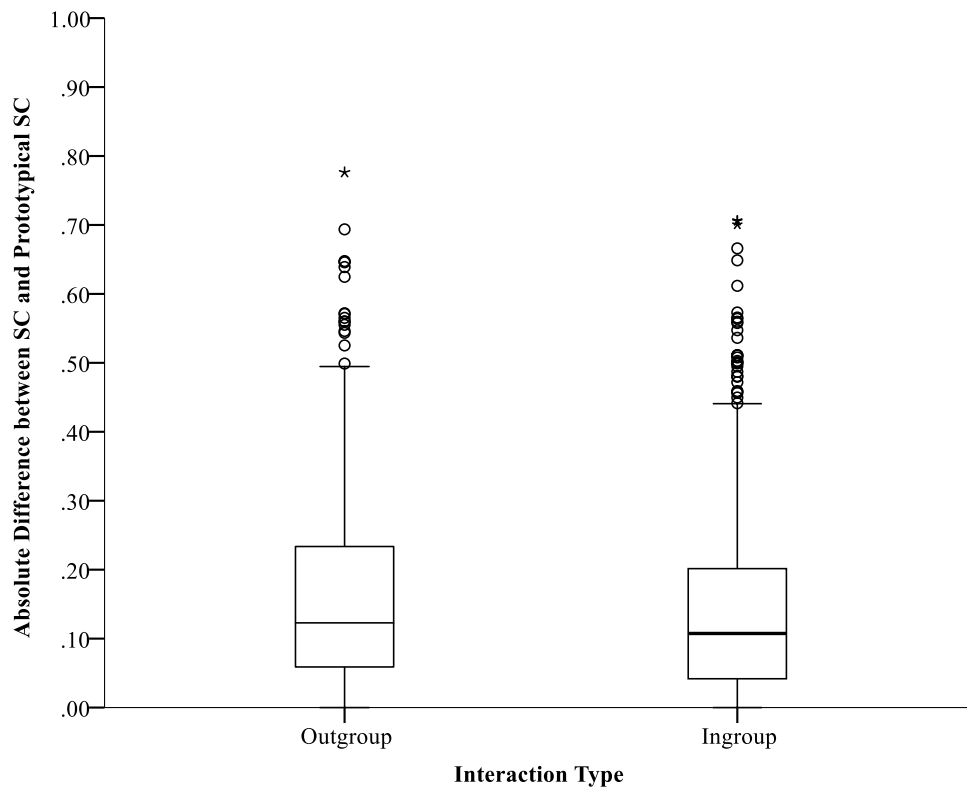


Figure 4.8. Polarisation of positions on violence in intergroup and intragroup interactions (Intergroup and Intragroup interactions each $N = 1000$)

After establishing significantly larger polarisations of opinions in intergroup interactions, we tested H2: the intergroup hostility bias. A one-way analysis of covariance (ANCOVA) was conducted to ascertain whether the PoV values in intergroup interactions were significantly higher than the PoV values in intragroup interactions when controlling for several covariates. These covariates included the dispositions of the two interactants and the received instigation. Notably, disposition and instigation were the ‘non-social’ terms of the aggression calculation and were included as covariates to isolate any potential ‘social’ effect. The last covariate included was the number of bystanders present when determining a prototype. There was no significant difference in the PoV values between intergroup interactions and intragroup interactions (estimated marginal means = 0.11 and 0.11 respectively; $F(1, 1995) = 0.82, p = .365, \eta^2 p < .001$). As we found no effect of interaction type on PoV values, we did not attempt to test whether this effect was mediated by the marginally larger polarisation found in intergroup interactions.

4.3.4 Discussion

In this study, we assessed whether the social model of violence (from Study 4) produced statistically higher levels of intergroup violence than intragroup violence. We expected higher intergroup violence, as individuals in intergroup contexts would be more likely to polarise their positions on violence away from the outgroup comparison, thus resulting in more extreme positions on violence (Borum & Pynchon, 1999; Viki & Abrams, 2013; M. A. Wilson et al., 2009).

As predicted, we found evidence that individuals in intergroup interactions polarised their opinion significantly more than those in intragroup interactions. However, this group interaction effect was weak, and on average only resulted in a 0.03 polarised shift on a scale from 0.00 to 1.00. Although we found statistical evidence of greater polarisation variation in intergroup interactions, examination of the descriptive box plots suggest that intragroup interactions also recorded a large level of variation between original positions on violence and prototypical positions adopted. Therefore, two questions remain: first, why are ingroup members not all reaching one unified position on violence close to an individual's starting position? Second, why are intergroup encounters (on average) only polarising marginally more than intergroup encounters? To address the question regarding narrow consensus of intragroup positions, it is important to refer back to Salzarulo's (2004) original polarisation mechanism. In an intragroup context with no 'outgroup presence', Salzarulo (2004) assumed the prototypical position of the group would be central and close to one's initial starting position¹⁸ (see Figure 4.9).

¹⁸ Though see Mackie (1986) and Mackie and Cooper (1984) for evidence that ingroup polarisations of norms and attitudes can occur in the absence of a comparative outgroup.

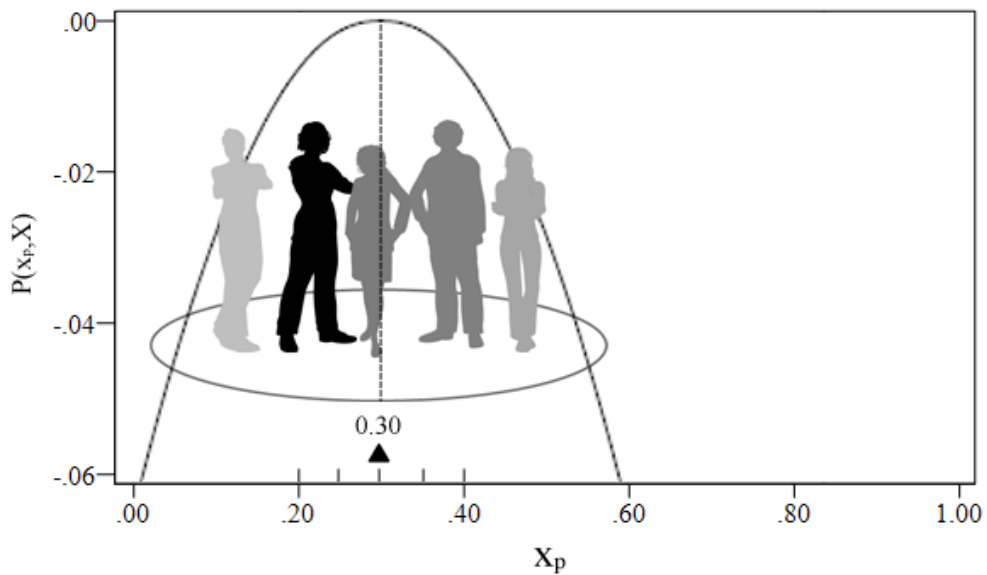


Figure 4.9. Colour social group of five agents with similar positions on violence.
Adapted from Salzarulo (2004)

However, even in the depicted example, individuals can still expect some deviation from their original position on violence (e.g., the absolute difference from agent 0.20 to the prototypical position of 0.30 is 0.10, which is fairly large).

However, in practice, it would not be expected for individuals within a social group to be so close on initial positions of violence (SC). As aforementioned (see discussion in Study 4), the social groups in this model form together on dimensions of redness and greenness, and not on positions of violence (SC). While the colour membership function means that an individual is more likely to take fellow social group members' positions on violence into account when deriving a prototypical position, there is no guarantee of consensus. If an individual has a very different prototypical position on violence to other colour social group members (see Figure 4.10), that individual may recategorise and polarise away from these fellow social group 'others' (in the comparative context of violence, rather than redness and greenness). This would be an example of large polarisation occurring within an isolated intragroup context.

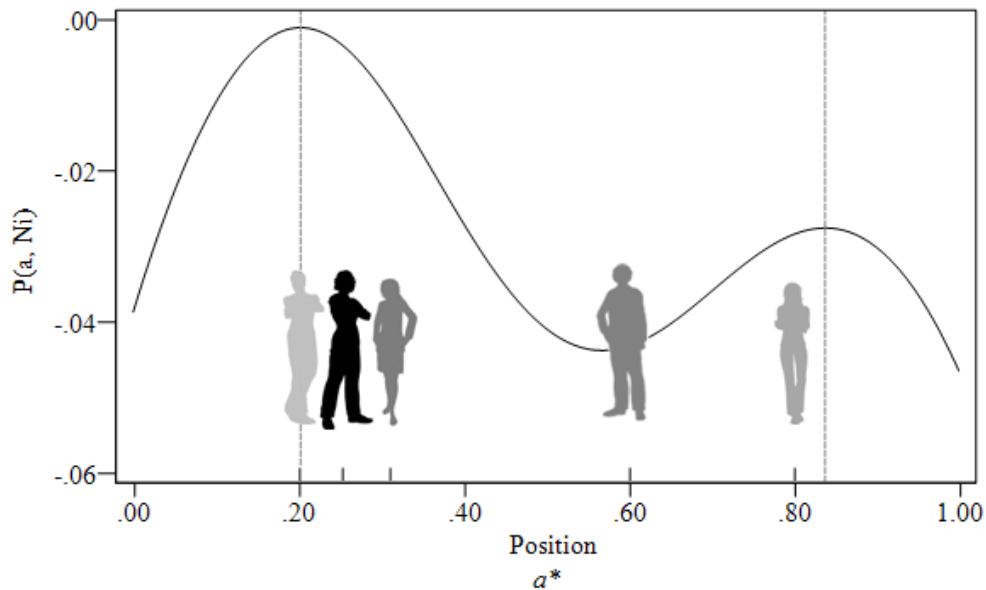


Figure 4.10. Colour social group of five agents with divergent positions on violence

It is also important to recall that agents considered the positions of *all* other agents within the local vicinity (their neighbourhood) when deriving a prototype (see step 3 in ‘Model Steps’). Following this step, an agent then selected *any* other agent within the local vicinity for a potentially aggressive interaction (see step 4 in ‘Model Steps’). As an accumulation, it is entirely possible that an ‘intragroup’ aggressive interaction could take place within the proximity of colour ‘outgroup’ members. Although classified as an ‘intragroup’ interaction, large polarisation may occur in response to the presence of these comparatively different colour outgroups. Alternatively, the presence of colour ‘outgroup’ members may present the opportunity for individuals engaging in an intragroup interaction to recategorise from their own social group members, and to reach a shared prototypical position on violence with colour ‘outgroup’ members more similar on the relevant dimension of violence. Again, this recategorisation may result in large levels of polarisation in an intragroup interaction. Hypothetically, in this scenario, the same agent could just as easily have randomly selected an ‘outgroup’ member to interact with as an ‘ingroup’ member (see step 4 in ‘Model Steps’). Although this selection would make no numerical difference to the prototypical position on violence (which has already been calculated by this stage from all

other agents within the local vicinity in step 3), it would, however, result in a different interaction value (0 instead of 1) being recorded in the results file. This roll-of-a-dice interaction value is then used to determine if there are statistical differences at the group level.

In short, an expectation of small polarisations in intragroup interactions and large polarisations in intergroup interactions was too simplified. In even the most isolated intragroup cases, individuals may still recategorise from fellow social group members and polarise away from them if their positions on violence deviate too much. Close proximity to outgroup members presents the opportunity for both higher polarisation from these ‘others’ and also recategorisation with these ‘others’ if their positions on violence better suit one’s own. Both of these examples, although distinct, may result in large polarisation of an individual away from their original position on violence (e.g., polarising with group members away from one’s original position in the presence of an outgroup, or recategorising with outgroup members and polarising away from one’s original position with them).

Our second prediction was that the PoV values derived from intergroup interactions would be significantly higher than the PoV values derived from intragroup interactions. We found no evidence to support this expectation, and, therefore, did not conduct further analyses to investigate if this was mediated by larger (albeit only marginally larger) levels of polarisation in intergroup encounters.

However, on further reflection, even if our model had produced large levels of polarisation in intergroup encounters and small levels of polarisation in intragroup interactions, the prediction that this polarisation would lead to group differences in levels of violence may be an illogical inconsistency (Galán et al., 2009). Polarisation from comparatively more law-abiding others has been assumed to explain increased levels of antisocial behaviour in lesser criminals (Viki & Abrams, 2013) and high levels of violence in the domain of extreme terrorist groups (Borum & Pynchon, 1999; M. A. Wilson et al., 2009). However, polarisation is exactly that: polarity. As such, a push towards a more extreme

position of violence (e.g., violence is acceptable) is met with an equal opposing push away towards the alternative position (e.g., violence is not acceptable). No matter how large the magnitude of the push is in either direction, the average position on violence would still meet in the middle. This same principle holds true for intragroup polarising encounters—even if the magnitude of the push is relatively smaller, the average position on violence would still meet in the middle. Therefore, unless there is an uneven number of individuals on each polar end (e.g., five individuals polarised from 0.3 to 0.1 becoming more susceptible to violence and only two counter individuals polarised from 0.7 to 0.9 becoming more peaceful), we would expect the average push position among agents over the course of the simulation to cancel one another out.

4.4 Study 5b: A Further Examination of Polarisation and the Effect on Violence

To examine the question of polarisation further, we manipulated a modelling scenario designed to create very small within-group polarisation and very large between-group polarisation. We then tested whether this significantly larger level of polarisation in intergroup encounters was related to higher levels of intergroup violent interactions.

4.4.1 Method

To ensure that all ingroup members had similar positions on violence (SCs), we instructed all agents to form groups based on the dimensions of greenness and positions of violence (SC), rather than redness and greenness. To guarantee that ingroup interactions would not be influenced by outgroup presence, we only recorded intragroup interactions that occurred when there were no outgroup agents present within an individual's neighbourhood. Finally, to ensure a push away from non-prototypical positions, we increased the aversion (α) parameter—where zero equals no repulsive effect due to outgroup—from Salzarulo's (2006) 0.08 calibration to the most extreme position in his paper, 0.30. All other model specifications

and procedures remained unchanged. Once again, we randomly selected 1000 intragroup interactions and 1000 intergroup interactions for data analysis.

4.4.2 Results

Figure 4.11 shows the absolute difference between an individual's starting position on violence (SC) and the final prototypical position (v_i^*) adopted after metacontrast calculations. As in the original study, there appears to be greater polarisation in intergroup interactions in comparison to intragroup interactions. Visually, the magnitude of difference between the two interaction types appears larger in this manipulated study (see Figure 4.11) than in the original study (see Figure 4.8).

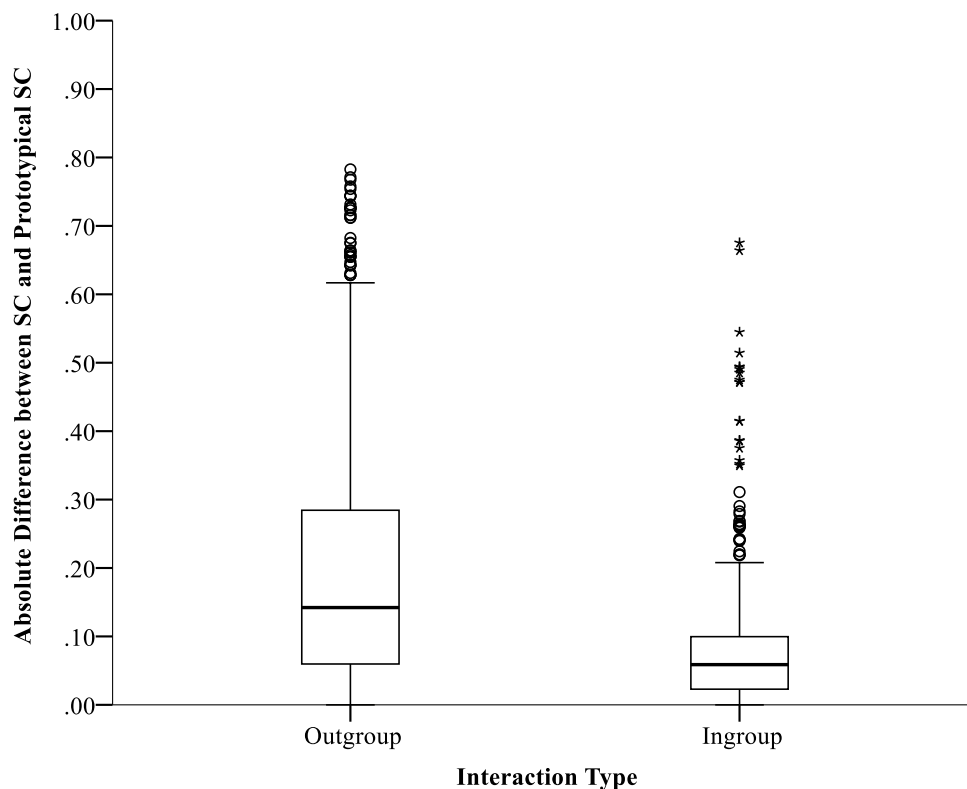


Figure 4.11. Polarisation of positions on violence in intergroup and intragroup interactions (intergroup and intragroup interactions each $N = 1000$)

A Wilcoxon signed-rank test for related samples again found that the difference in average levels of polarisation statistically differed between interaction types (average intergroup polarisation: $M = 0.20$, $SD = 0.18$, average intragroup polarisation: $M = 0.07$,

$SD = 0.08$; $Z = -18.19$, $p < .001$, $r = -.51$). This large effect size was far greater than the 0.07 weak effect reported in the original experiment. Akin to the original experiment, we found a statistically greater divergence in polarisation values for intergroup interactions compared with intragroup interactions (Fligner-Killeen's test of homogeneity of variances, $\chi^2(1, N = 2000) = 506.9$, $p < .001$). From this information, we can conclude that our model manipulations were successful in creating scenarios with small within-group polarisations and large between-group polarisations.

In order to appropriately depict how polarisation is bidirectional with an overall central average, we plotted the difference between an individual's original position on violence (SC) minus their final prototypical position (v_i^*) adopted after metacontrast calculations (see Figure 4.12). Again, intergroup interactions displayed a wider magnitude of polarisation, but the medians between interaction types appeared almost identical. A Wilcoxon signed-rank test for related samples found the mean value between an individual's original position on violence (SC) and their final prototypical position (v_i^*) adopted after metacontrast calculations to be non-statistically different at the interaction level (average intergroup difference: $M = 0.008$, $SD = 0.27$; average intragroup difference: $M = 0.005$, $SD = 0.11$; $Z = .31$, $p = .754$, $r = -.007$). In short, although there is greater polarisation in intergroup encounters, the overall average position on violence is the same for intergroup encounters and intragroup encounters, with the means around zero in both instances.

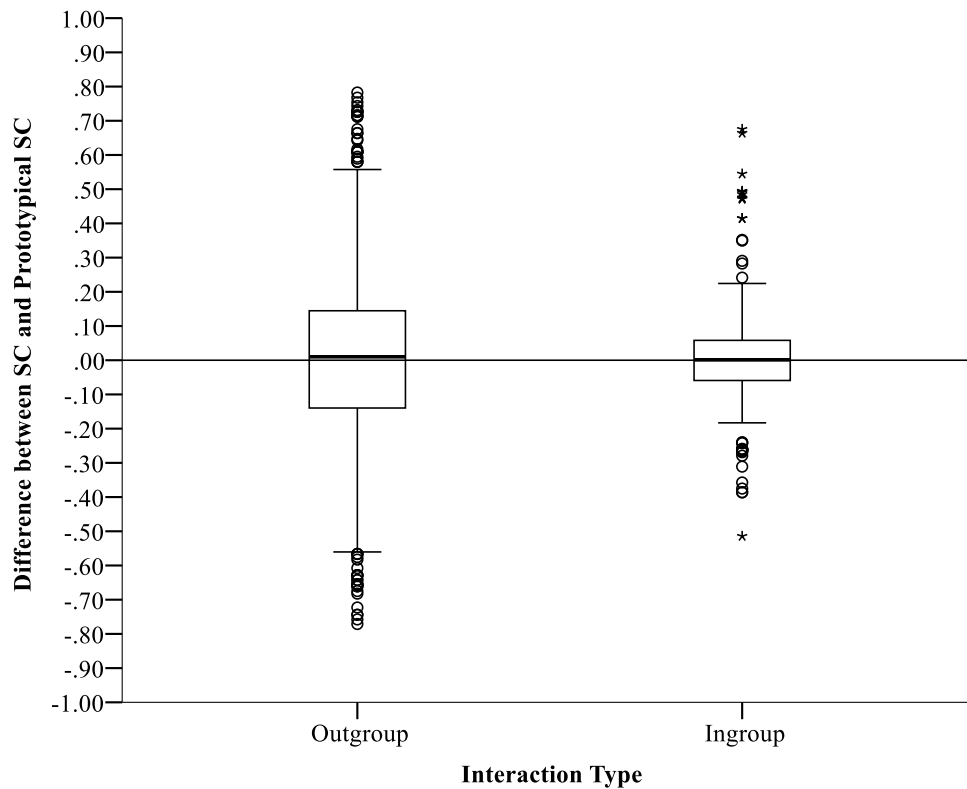


Figure 4.12. Bidirectional polarisation of positions on violence in intergroup and intragroup interactions (intergroup and intragroup interactions each $N = 1000$)

A subsequent one-way ANCOVA found no statistical difference in the PoV values between intergroup interactions and intragroup interactions when controlling for the dispositions of the interactants and the received instigation (estimated marginal means = 0.14 and 0.14 respectively, $F(1, 1995) = .56, p = .452, \eta^2 p < .001$). Further, there was no evidence of a relationship between the absolute distance polarised from one's original position and the PoV (Pearson's $r = -.004, N = 2000, p = .864$). These results support the idea that larger polarisations in intergroup encounters cannot be expected to drive overall higher levels of intergroup hostility. It further suggests that without further modifications, an intergroup hostility bias is unlikely to be an emergent property of the current social model of violence.

4.4.3 Discussion and ABM Evaluation

Agent-based modelling allows us to dynamically compare a dyadic theoretical model of aggression (adapted from I³ theory) with an alternative theoretical model that also

incorporates further dynamics between actors who share prototypical positions on violence. In comparing these two worlds, we largely found inhibition and disinhibition of violence from the local context and social groups in the social model, but less so in the dyadic model. We do not claim that this social model is an improvement on existing models; however, it is an interesting alternative in which small social considerations can make important qualitative and quantitative differences.

One of the challenges posed in this modelling process, was how to analyse data that dynamically unfolds and alters over time? Typically, it is possible to test group-based trajectories over time using statistical methods such as repeated measures MANOVAs or Multi-level modelling. However, to use these conventional statistical approaches the groups need to be consistent over time. Throughout the simulation, our agents changed their allegiances to different social groups in response to changes in the comparative context. This made conventional statistical methods inappropriate for our purposes. One way to circumvent this challenge could be to keep group assignments constant throughout the course of the simulation. However, this would incorporate temporal dynamical testing at the expense of fluid social group membership. As we were interested in the dynamics of social group membership, and how identities can emerge and change, our statistical techniques for this modelling work were restricted to snapshots of time.

When considering the results of our studies, we found no evidence of an emergent intergroup hostility bias (as found in our CCTV analysis, Study 1) from the social model. Even when adjusting the metacontrast parameters to increase the potential for polarisation in intergroup interactions, and when instructing groups to form on positions of violence, no such bias emerged. This is because polarisation works in both directions, resulting in an unchanged prototypical average.

By failing to find the hostility bias evidenced in real-world footage, it could be suggested that the model provides a poor fit of reality. Alternatively, one could propose that

the social model reflects reality well, and that CCTV footage sampling may be prejudicing our expectations towards an intergroup hostility bias. Specifically, CCTV operatives were asked to identify disputes that looked like they were escalating. If a group in a public space becomes polarised towards peace and exits a situation promptly, this interaction is less likely to be captured by the CCTV operatives. As such, it could be the case that the selection bias inherent in CCTV footage cuts out the peaceful tail end of polarisation, leaving only the more hostile events for future analyses. An alternative explanation for the absence of intra-intergroup differences in levels of violence could be explained by the model's lack of alternative theoretical processes that may reduce intra-group violence (such as the expectation of future intragroup interactions; de Waal, 2000) and inter-group violence (such as superordinate frameworks and norms that encourage 'fair fights' amongst all participants; see Copes, Hochstetler and Forsyth, 2013).

Alternatively, if a researcher disputes this claim and anticipates that a metacontrast informed model of group violence alone *should* produce an intergroup hostility bias, then it is important to affirm that we cannot expect such a bias to reliably arise from this social model without further modifications. Reproduction of an intergroup bias can be easily achieved by explicitly putting the bias into the model. For example, one could increase the instigation (provocation) using a multiplier (β) that is larger for intergroup encounters than intragroup encounters, so that provocations are modelled as more inflammatory if coming from an outgroup member. This is incorporated into the model as:

$$Aggression\ Calculation = \frac{\beta \times Instigation}{Disposition + Prototypical\ Social\ Component}$$

where $\beta = 1$ for intragroup encounters and $\beta > 1$ for intergroup encounters. This is in line with the impression management literature whereby one is presumed to react more strongly to an insult from another group (Felson, 1978). It is also in line with the theoretical assumptions of Otten (2009) and the behavioural analyses of Liebst, Heinskou and Ejbye-Ernst (2017) who

proposed that social categorisations should matter in aggressive interactions in so much that more negative connotations are perceived and reacted to when the other is ‘one of them’ as opposed to ‘one of us’. This hostility bias modification yields the desired effects (separately tested but not reported here), but it is a result that depends on explicitly instructing agents to hold a hostility bias when encountering (and being provoked by) outgroup agents. Altering the model in this way does not produce emergent behaviour, but provides a ‘what you put in is what you get out’ result. As such, it ‘reproduces’ the bias found in the CCTV footage, rather than calibrating a model towards the CCTV data. This forced entry approach runs counter to the emergent procedure adopted in Studies 3–5b, whereby we figuratively ‘planted the seeds’ (Epstein, 1999) with no full expectation of what might grow.

On this point, it is important to step back from the minutiae of this specific model and reconsider the wider objectives of the thesis. Specifically, the social ABM was designed as a theoretical proof of concept to compare dyadic models with an alternative that had shared prototypical positions and local normative behaviour. If satisfied that an emergent model of norm formation offers a nice theoretical alternative to the I³ model with some interesting results, then there may be no gain in purposely manipulating the theoretical model to obtain the result expected from CCTV footage study. Further, if our ultimate objective is to create an ABM derived from the CCTV findings that can allow us to make some important predictions about how people may behave in public space violence, calibrated against the CCTV data available, then we have to be realistic about the limitations of our theoretical model.

Specifically, the theoretical social model acts as a cognitive model in which agents are permitted to determine how they ought to behave based on the agents and social context around them. However, the CCTV footage does not allow us to examine cognitive processes of individuals nor norm formation and deployment. Rather, the footage only allows us to observe what people actually do. We then infer what is normative from the behavioural

distinctions observed, for example, the typical response to ingroup escalation is ingroup de-escalation; therefore, it is normative to de-escalate ingroup members.

As such, to make *behavioural predictions* of public space violence (e.g., how group membership may affect rates of intervention, the importance of third-party presence and the sequence of interactions and reinteractions that lead to violence escalation and de-escalation), we feel it is important to return to the CCTV data for a more intricate secondary data analysis. This refined information can then help furnish future behavioural (as opposed to theoretical) ABMs.

In the following chapter, we will extract more refined information from secondary microanalyses of our CCTV data. We will:

- i) explore ingroup and outgroup behavioural differences when unpacking escalation into physical and non-physical behaviours
- ii) conduct an additional sequence analysis of what happens next after an escalatory or de-escalatory action within and between groups
- iii) examine the standardized number of interactions typically performed by protagonists, targets and bystanders to ascertain relative level of involvement
- iv) dissect the different bystander roles enacted by third-parties and propose a new typology of intervention based on pacifying and policing roles.

5 A Secondary Re-Examination of Intergroup Hostilities and the Role of Third-Party Involvement in NTE Violence

In this chapter, we refine our understanding of intergroup / intragroup antagonistic behavioural distinctions with a secondary data analysis of the CCTV footage coded in Study 1. We investigate the relative level of bystander involvement in episodes of NTE violence and examine whether third-party intervention is a division of labour with degrees of specialisation that can impact on violence severity. An objective of this extended work is to provide more fine-grained information extracted from the footage, which can then be used to inform future integrated CCTV behavioural ABMs of public space group violence.

In the first study of this chapter (Study 6), we extend Study 1 and examine intragroup/intergroup behavioural differences when unpacking escalation into physical and non-physical behaviours. We re-examine the behavioural sequence of actions from Study 1 and assess which behaviours typically succeed acts of intragroup/intergroup physical and non-physical escalation. We inspect the relationship between physical escalation and the number of bystanders, and the relationship between non-physical escalation and the number of bystanders, for both intragroup and intergroup videos. Finally, we determine the average level of third-party bystander involvement. To this end, we compare the relative number of actions contributed by bystanders to the relative number of actions performed by protagonists and targets.

In the second study of this chapter (Study 7), we further dissect the different third-party bystander roles and propose a new typology of intervention based on pacifying and policing roles. Specifically, we examine the distinctions between ‘active bystanders’ who do not address the protagonist and target; ‘pacifiers’ who perform de-escalatory actions towards one of the opponents; and ‘policers’ who perform de-escalatory actions towards both of the opponents. We summarise how a new typology of bystander roles may inform violence intervention initiatives.

5.1 Study 6: An Extended CCTV Microanalysis of Within- and Between-Group Aggressive Acts and Level of Bystander Participation

In Study 1, we demonstrated an intergroup hostility bias, whereby a significantly larger proportion of intergroup-directed actions were escalatory in comparison to intragroup directed actions. We supported these findings with literature suggesting that categorisations of ‘us’ versus ‘them’ *during conflict* is strongly related to derogation towards the opposing ‘others’ (Decker, 1996; Liebst et al., 2017; Struch & Schwartz, 1989; Tedeschi & Felson, 1994; Vasquez et al., 2015). In contrast, our agent-based modelling work with integrated metacontrast principles (Studies 5a and 5b) found no evidence of an overall intergroup hostility bias. This was because groups in our modelling environment polarised towards both violence *and* peace. However, in our studies above, we did not unpack escalation into physical and non-physical behaviours. Further assessment of these distinct categories could enhance our understanding of the intergroup / intragroup hostility behavioural differences demonstrated in Study 1. Extraction of this more fine-grained information may further aid the development of an integrated behavioural ABM of violence calibrated against the actual CCTV data. Therefore, the first aim of this study is to assess whether there exists an intergroup physical escalatory bias and an intergroup non-physical escalatory bias in episodes of public space violence.

In Study 1, we ascertained that third-parties were active players during public space conflicts, and contributed 66.3 % of the total actions recorded. However, a limitation of this study was that we did not examine the *relative* number of actions performed by each role: protagonist, target and active bystander. Therefore, to establish the relative involvement levels of third-parties, we also report the standardised number of actions contributed by each role: protagonists, targets and bystanders.

5.1.1 Hypotheses

Although this study is exploratory—to refine our understanding of intragroup / intergroup behavioural differences and to provide more fine-grained information extracted from the footage for future modelling purposes—we do provide several theoretical predictions.

In Study 1, we found a hostility bias, whereby intergroup interactions during conflicts were more likely to be hostile than intragroup interactions during conflicts. In contrast, our agent-based modelling work (Studies 5a and 5b) found no evidence of an intergroup hostility bias. Therefore, we could predict *either* comparably higher levels of physical and non-physical escalation in intergroup interactions (as found in Study 1) *or* a null finding (in line with Studies 5a and 5b). In our Chapter 4 discussion, we put forward the proposition that the CCTV sampling method itself may be prejudicing our expectations of an intergroup hostility bias. Specifically, that CCTV operatives by capturing disputes that look like escalating, may miss the intergroup conflicts in which a group polarises towards nonviolence and exits a situation without gaining attention. However, as the current study is a secondary data analysis from the CCTV work coded in Study 1, we may again expect an intergroup hostility bias when separating escalation into physical and non-physical escalation. Therefore, we predict:

- H1:** There will be a significantly larger proportion of intergroup physical escalatory actions than intragroup physical escalatory actions.
- H2:** There will be a significantly larger proportion of intergroup non-physical escalatory actions than intragroup non-physical escalatory actions.

Regarding the standardised number of actions contributed by each role, as the protagonists and targets are the actors initially involved in the conflict, and as the protagonist and target are central to most theories of aggression (e.g., I³ theory; Finkel, 2007, 2014), we predict:

H3: That protagonists (H3a) and targets (H3b) will, on average, contribute significantly more acts than will bystanders.

As protagonists performed significantly more actions than targets in Study 1, and as there is an equal number of protagonists and targets across videos, we predict:

H4: That protagonists on average will contribute significantly more acts than will targets.

5.1.2 Methods

As these were secondary data analyses, the procedures around data collection and video coding remain the same as Study 1. To separately assess physical escalation and non-physical escalation, it is important to subdivide the ‘escalation’ category from Study 1. As the original coding distinguished between non-physical escalation and physical escalation, no further procedure was required to separate the two. We were also interested in the relative number of actions performed by each role: protagonist, target and active bystander. As each video varied in length, number of actions and number of individuals, we quantified participation as the standardised number of actions for each participant within a video. These standardised figures were *z* scores: the distance between the number of actions performed by an individual and the mean number of actions performed by all individuals within the same video. This allowed comparison of the average standardised number of actions for each role; that is, did one role perform comparatively more actions over the course of the 43 videos than another role?

5.1.3 Results

To briefly reiterate the descriptive statistics of Study 1, the majority of the 3555 actions coded across the corpus of 43 CCTV were de-escalatory in nature (e.g., open hand calming gestures, de-escalatory touching and movements to block contact, etc.). The remaining 36.9 % (1312) of behaviours were escalatory.

Subdivision of the escalation category into physical escalation and non-physical escalation showed that the occurrence of each behaviour was very similar (Figure 5.1). Fifty percent (657 out of 1312 acts) were *physical escalation* (e.g., kicking, punching, slapping, grappling, pushing, etc.) and 50 % (655 out of 1312 acts) were *non-physical escalation* (e.g., aggressive gesturing, pointing, invading of other's personal space, removing of one's own clothes, etc.); $\chi^2(1, N = 1312) < .01, p = .956$, Cramer's $V = .002, n.s.$

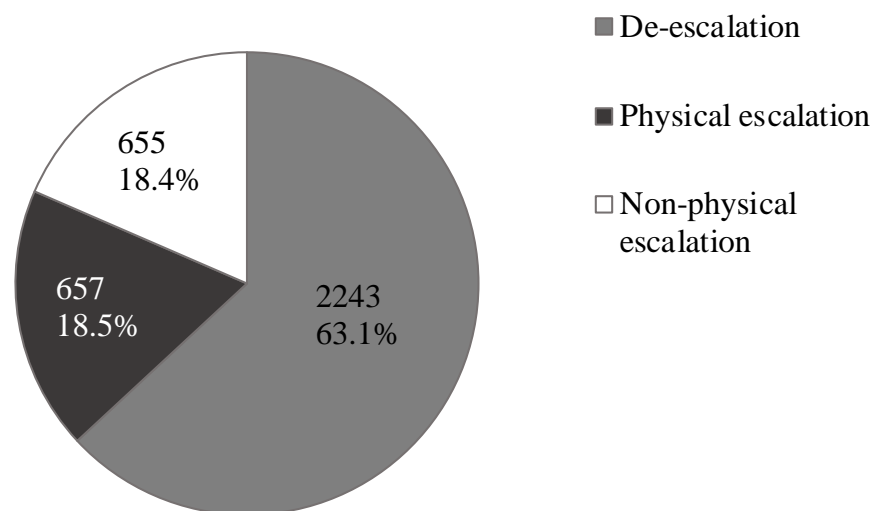


Figure 5.1. Proportions of de-escalatory, physical escalatory and non-physical escalatory acts as a percentage of the sum of all behaviours, $N = 3555$

To again account for the potential that participants on average across all videos had greater exposure to ingroup members than did outgroup members, and thus potentially a higher likelihood of intragroup interaction, and to account for variations in the lengths of videos, the total number of intragroup and intergroup escalatory actions (physical escalation and non-physical escalation) for each video were converted into proportions. Intragroup behavioural proportions were attained by dividing the number of intragroup physical escalatory acts and intragroup non-physical escalatory acts in each of the 43 videos by the total number of intragroup acts in that video. In similar fashion, intergroup behavioural proportions were attained by individually dividing the number of intergroup physical escalatory acts and intergroup non-physical escalatory acts in each of the 31 videos (containing intergroup behaviours) by the total number of intergroup acts in that video.

To identify general trends in the proportional compositions of de-escalation, physical escalation and non-physical escalation across videos and group condition, these data were visualised in a barycentric coordinates systems plot (Möbius, 1827, as cited in Coxeter, 1969), which graphically depicted the proportions of the three behaviours as points in an equilateral triangle. In this plot (see Figure 5.2), each coordinate point represented a video: black points for intragroup videos ($N = 43$) and white for intergroup videos ($N = 31$). Each coordinate point is positioned according to a triple number that signified the proportion ratio of physical escalation, non-physical escalation and de-escalation for that video. For example, video one, a black (intragroup) data point, had a coordinate triple of [0.15, 0.03, 0.82] representing that 15 % of the intergroup actions for that video were physical escalatory, 3 % were non-physical escalatory, and 82 % were de-escalatory. The mean coordinates across the 43 intragroup videos were represented by a '+' while the mean coordinates across the 31 intergroup videos were represented by an 'x'.

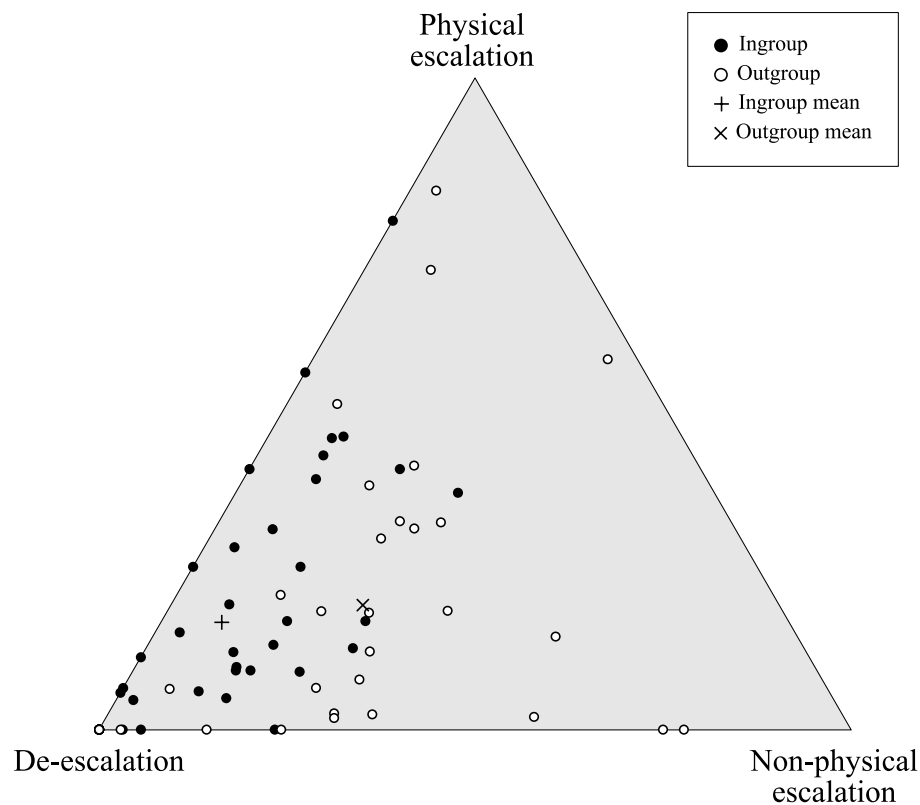


Figure 5.2. A barycentric coordinate systems plot representing the proportion ratio of physical escalation, non-physical escalation and de-escalation for each video, $N = 74$

Initial inspection of the scatter distribution between points showed large variation in the proportion of actions between cases—though this scatter appeared larger for intergroup cases than intragroup cases. Intragroup cases appeared to be distributed along the de-escalation/physical escalation axis, while intergroup cases appeared to be distributed more centrally and along the de-escalation/non-physical escalation axis. There were few cases with large proportions of both physical escalation and non-physical escalation and these occurred only for intergroup interactions. All cases with a zero proportion of non-physical escalation were taken from the intragroup sample. Across the majority of cases, de-escalation was the largest proportion of actions, though the scatter appeared to be more heavily distributed towards de-escalation in intragroup cases than intergroup cases (as already statistically confirmed in Study 1). The scatter distribution generally showed a higher proportion of non-physical escalatory actions in intergroup cases than in intragroup cases. The scatter distribution for physical escalation appeared similar for both intragroup and intergroup cases.

These general deductions are reflected in the mean markings ('+' and 'x'), which showed, on average, similar levels of physical escalation at the group level (16.5 % of all intragroup actions vs. 19.1 % of all intergroup actions), but higher levels of intragroup de-escalation than intergroup de-escalation (75.4 % vs. 55.4 % respectively), and lower levels of intragroup non-physical escalation than intergroup non-physical escalation (8.1 % vs. 25.5 %). Put simply, there appeared to be a proportionally higher level of de-escalatory actions within groups than between groups (as statistically confirmed in Study 1), and a proportionally higher level of non-physical and physical escalatory actions between groups than within groups (see Figures 5.3 and 5.4). However, these proportions were not weighted by the total number of actions.

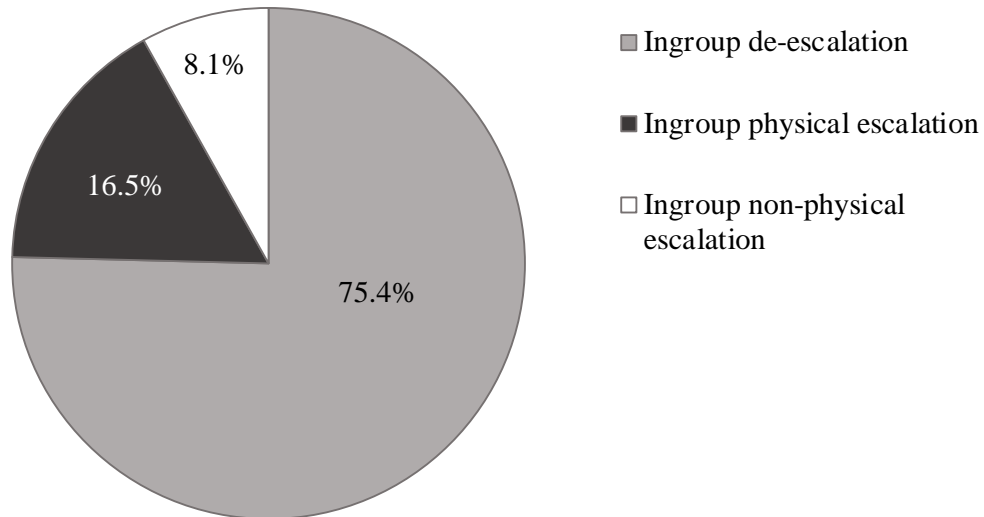


Figure 5.3. The average proportion of intragroup de-escalation, physical escalation and non-physical escalation across the 43 cases

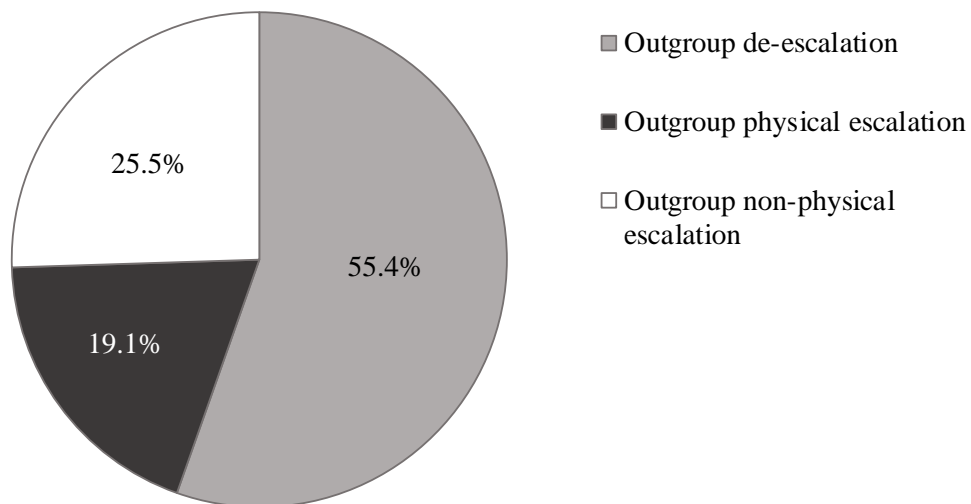


Figure 5.4. The average proportion of intergroup de-escalation, physical escalation and non-physical escalation across the 31 cases

Two binomial logistic regressions of proportions (Warton & Hui, 2011) were conducted to ascertain whether the average weighted¹⁹ proportions of physical escalation and non-physical escalation differed statistically at the group level when controlling for several covariates. Akin with Study 1, these covariates included number of active bystanders, the severity of violence and the type of conflict captured (whether primarily an intragroup conflict

¹⁹ To retain information on a sample size, all proportions were weighted by the total number of acts from which the proportion was estimated.

or intergroup conflict). In addition to these covariates, both logistic regression models contained a random intercept term to account for overdispersion (see Warton & Hui, 2011). As the proportions of de-escalation remained unchanged from Study 1, we did not repeat a binomial logistic regression on this proportion.

We first tested Hypothesis 1, whether there is a significantly larger proportion of intergroup physical escalatory actions than intragroup physical escalatory actions. The first binomial logistic regression comparing the physical escalatory proportions of intragroup actions and intergroup actions when controlling for covariates was not statistically significant, $\chi^2(6, N = 74) = 0.95$, 95% CI [-0.98, 0.26], $p = .331$, *n.s.*, odds ratio = 0.76. Therefore, intergroup actions did not contain statistically higher proportions of physical escalation than intragroup actions.

We next tested Hypothesis 2, whether there is a significantly larger proportion of intergroup non-physical escalatory actions than intragroup non-physical escalatory actions. The second binomial logistic regression comparing the non-physical escalatory proportions of intragroup actions and intergroup actions when controlling for covariates was statistically significant, $\chi^2(6, N = 74) = 27.43$, 95% CI [-2.16, -1.43], $p < .001$, odds ratio = 0.20. Intergroup action cases contained a significantly higher proportion of non-physical escalation than intragroup action cases.

In sum, when de-aggregating escalation into subcategories, there was an effect of group on the proportion of non-physical escalatory acts: intragroup interactions contained proportionally less non-physical escalation than intergroup interactions. No effect of group was found on the proportion of physical escalatory acts—intragroup interactions were proportionally as likely to be physically violent as intergroup interactions.

Examination of the action sequence (Table 5.1) showed some interesting behavioural tendencies following intragroup-directed escalatory actions and intergroup-directed escalatory actions. After an act of intragroup physical escalation, 47.8 % of the subsequent actions were

intragroup de-escalation and 37.9 % of the subsequent actions were intragroup physical escalation. This difference was statistically significant ($Z = 2.67$, $SE = 0.04$, $p < .001$). In contrast, after an intergroup physical escalatory action, 48.5 % of the subsequent actions were intergroup physical escalation and 28.8 % of the next actions were intergroup de-escalation ($Z = 4.95$, $SE = 0.04$, $p < .001$). This represented a behavioural reversal in the most predominant response to physical escalation by interaction type—that is, predominately a de-escalatory response to intragroup directed physical escalation and predominantly a physically harmful response to intergroup directed physical escalation. The differences in these proportions were statistically significant (intragroup de-escalation following intragroup physical escalation 47.8 % vs. intergroup de-escalation following intergroup physical escalation 28.8 %, $Z = 4.96$, $SE = 0.04$, $p < .001$; intragroup physical escalation following intragroup physical escalation 37.9 % vs. intergroup physical escalation following intergroup physical escalation 48.5 %, $Z = -2.73$, $SE = 0.04$, $p = .006$). As such, physical escalatory acts tended to be responded to differently contingent on whether they were directed within groups or between groups.

After an act of intragroup non-physical escalation, the predominant following act was intragroup de-escalation (58.9 %), with the next likely response intragroup non-physical escalation (18.5 %, $Z = 9.24$, $SE = 0.04$, $p < .001$). A similar pattern was found in the case of intergroup non-physical escalation. After an intergroup non-physical escalatory act, the predominant following act was intergroup de-escalation (33.0 %), with the next most likely response intergroup non-physical escalation (30.0 %). However, there was no statistical difference between the level of subsequent intergroup de-escalation (33.0%) and level of subsequent intergroup non-physical escalation (30.0 %, $Z = -0.92$, $SE = 0.03$, $p = .36$, *n.s.*). As such, intergroup non-physical escalation was statistically as likely to be met with further intergroup non-physical escalation as intergroup de-escalation. However, it is important to note that intergroup non-physical escalation was met with intragroup de-escalation on 24.3%

of occasions. Therefore, the most likely aggregated response to intergroup non-physical aggression (intergroup de-escalation 33.0 % and intragroup de-escalation 24.3 % combined) was de-escalation—57.3 % vs. 42.7 % escalation ($Z = 4.15$, $SE = 0.04$, $p < .001$). This de-escalation, it seems, could occur across the group boundaries, or from self-policing within the group.

Table 5.1

The likelihood of the next sequential action succeeding an act of intragroup/intergroup physical escalation, intragroup / intergroup non-physical escalation and intragroup/intergroup de-escalation, N = 3511

Action	Next Intragroup Action			Next Intergroup Action			Sum (N)	
	Physical Escalation	Non-Physical Escalation	De- escalation	Physical Escalation	Non-Physical Escalation	De- escalation		
Intra- group	Physical Escalation	37.9 % (135)	5.6 % (20)	47.8 % (170)	0.08 % (3)	0.03 % (1)	7.6 % (27)	100 % (356)
	Non-Physical Escalation	10.1 % (25)	18.5 % (46)	58.9 % (146)	0.8 % (2)	2.8 % (7)	8.9 % (22)	100 % (248)
	De-escalation	10.6 % (152)	10.5 % (151)	57.7 % (830)	1.9 % (27)	7.9 % (114)	11.5 % (165)	100 % (1439)
Inter- group	Physical Escalation	0.7 % (2)	0.7 % (2)	13.4 % (40)	48.5% (145)	8.0% (24)	28.8 % (86)	100 % (299)
	Non-Physical Escalation	1.2 % (5)	1.7 % (7)	24.3 % (98)	9.7 % (39)	30.0 % (121)	33.0 % (133)	100 % (403)
	De-escalation	4.0 % (31)	2.5 % (19)	21.7 % (166)	10.4 % (80)	17.0 % (130)	44.4 % (340)	100 % (766)

Relationship between Number of Bystanders and Behaviours. Akin with Study 1, partial correlations were performed to determine the relationship between the number of active bystanders and the number of de-escalatory, physical escalatory and non-physical escalatory acts in a video while controlling for video length.

For videos depicting intergroup fights, there was no correlation between the number of bystanders and the number of physical escalatory acts ($r(20) = -.29, N = 23, p = .191, n.s.$) when controlling for these variables. However, there was a moderate, positive partial correlation between the number of bystanders and the number of non-physical escalatory acts (e.g., aggressive gesturing, posturing, pointing, invading of personal space and removing of own clothes, etc.), $r(20) = .43, N = 23, p = .048$.

For videos depicting intragroup fights, there was a statistically significant, moderate negative partial correlation between the number of bystanders and the number of physical escalatory acts: $r(17) = -.47, N = 20, p = .040$. There was no correlation between the number of bystanders and the number of non-physical escalatory acts when controlling for video length ($r(17) = -.44, N = 20, p = .058, n.s.$), though there was a non-significant negative relationship that was approaching statistical significance.

We next tested whether targets (H3a) and protagonists (H3b), on average, contributed significantly more acts than bystanders. We further tested whether protagonists (H4), on average, contributed significantly more acts than targets. To this end, we examined differences in the standardised number of actions contributed by each role: protagonists, targets and bystanders. Standardised z scores between roles showed that on average, protagonists performed comparatively more actions ($M = 0.62, SD = 0.81$) over the course of 43 videos than did targets and bystanders ($M = -0.08, SD = 0.93$ and $M = 0.007, SD = 1.00$, respectively). Each role had a large amount of variation, suggesting that some individuals performed relatively fewer / more actions than the average for that role (see Figure 5.5).

Bystanders had the largest variation in number of actions, and were responsible for providing both the most actions across all videos, and the least actions across all videos.

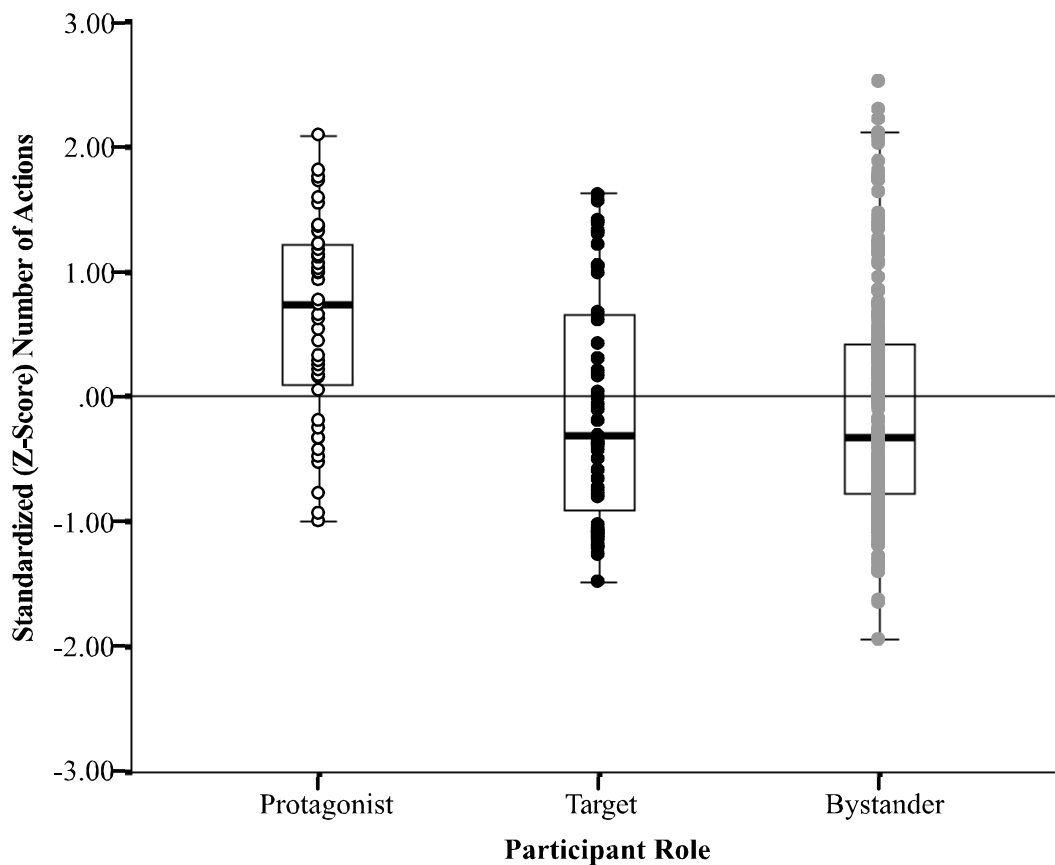


Figure 5.5. The standardised number of actions by role, N protagonists = 43, N targets = 43, N active bystanders = 242

Two Wilcoxon signed-rank tests for related samples found protagonists, on average, performed statistically more actions than targets ($Z = -3.43$, $p = .001$, $r = -.37$) and bystanders ($Z = -2.92$, $p = .003$, $r = -.17$). An additional Wilcoxon signed-rank test found no statistical difference between the average number of actions performed by targets and the average number of actions performed by bystanders ($Z = -0.41$, $p = .681$, *n.s.*, $r = -.02$).

5.1.4 Discussion

The purpose of this study was to perform a more fine-grained secondary analysis of the behaviourally coded video data, to help furnish future behavioural ABMs of public space violence. A further objective was to refine our understanding of intra/intergroup antagonist behavioural differences in episodes of public space violence. We re-examined the intergroup

hostility bias when de-aggregating escalatory behaviours into both physical escalation and non-physical escalation. Further, we reviewed the involvement levels of different roles: protagonist, target and bystander.

Consistent with the intergroup hostility bias and our predictions (H2), we found a significantly larger proportion of intergroup directed acts were non-physical escalatory in comparison to intragroup directed acts. However, we found no statistical difference in the proportion of intergroup physical escalatory acts and intragroup physical escalatory acts (H1). This similar likelihood of physical violence within group and between groups in episodes of public space violence is unexpected given the previously reviewed research. Specifically, the social categorisation literature predicts the existence of a positive ingroup bias, whereby insults and offences from outgroup members tend to be less likely pardoned than the equivalents from ingroup members (Otten, 2009). Social categorisation also predicts that individuals have a predisposition to feel concern for, and to support, their own group members by default over others (Reicher & Haslam, 2009) particularly during emergencies (Levine & Manning, 2013; Slater et al., 2013). This means that ingroup individuals may be more likely to de-escalate their own ingroup members during intra-conflicts, and be more willing to engage in partisan actions that support their own group members during inter-conflicts. This partisan support may be augmented in physical fights, in which there is a normative expectation that one should ‘stick up for their own’, even if that individual is in the wrong (Decker, 1996; Graham & Wells, 2003; Levine et al., 2012).

However, once again, an important distinction may need to be established between social category groups (Turner et al., 1987) and familiarity groups (Prentice et al., 1994; Swann et al., 2012). It is expected that the majority of groups who frequent the NTE (and who are thus captured in our footage) are friendship groups (Levine et al., 2012; Liebst et al., 2017). What makes friendship groups distinct from social category groups is the greater overall likelihood of familiarity and pre-existing social ties (de Waal, 2000; Swann et al.,

2012). On first consideration, increased familiarity, social ties and communal interdependence should encourage more cooperative behaviour and a greater motivation to stifle intragroup conflict before it can escalate into violence (de Waal, 2000; Krakauer et al., 2011). However, an alternative hypothesis is that familiarity increases the likelihood of intragroup conflict. For example, Levine, Lowe and Best (2012) interviewed 53 patrons of the NTE and found that interviewees were more willing to engage in violent exchanges with other friends than they were with strangers. For these interviewees, familiarity with friends equated to familiarity with the potential risks of conflict engagement. A familiar risk was more attractive than the unfamiliar risk from a stranger, which could result in serious injury (see Winlow & Hall, 2006 for similar reports that the potential of stranger violence generated more anxiety among interviewees than did the potential of friend violence).

Beyond familiarity of fighting capacity, Levine and colleagues' (2012) interviewees also provided insight into the acceptance of ingroup versus outgroup provocations. Specifically, interviewees specified that they were more willing to accept nonspecific provocations—such as the insult 'dickhead'—from friends over strangers (as predicted by Otten, 2009). However, familiarity with friendship group opponents could facilitate the construction of particularly intimate and degrading insults that could cause the familiar other to truly 'snap' (Levine et al., 2012). In more recent interview work on party drinking cultures, social closeness has been found to promote horseplay between familiar others (Pedersen, Copes, & Sandberg, 2016). This horseplay, which can start off in good spirits, may 'get out of hand' and provoke subsequent intragroup violence.

Tentatively, once again when considering risk, there may be an implicit expectation among friendship fighters that any quarrel will likely be broken up by others within the social group. The results from our sequence analysis largely support this assumption, and show that on average, almost half of the succeeding actions after physical escalation were ingroup de-escalation. Although intragroup conflict has alternative risks associated with damaged social

ties and weakened future cooperation (de Waal, 2000), these relationship considerations may be less problematic in the context of NTE violence where friends can “have a quick punch and then about half an hour later they are having a drink together” (Levine et al., 2012, p. 927). There may also be an expectation that fights within group should stay within group, and thus are less likely to receive police attention (Goudriaan & Nieuwbeerta, 2007). These factors combined offer the alternative proposition that social closeness may in fact facilitate an ingroup physical escalatory hostility bias, which may exist to counteract the outgroup physical escalatory bias so dominant within the literature. However, this ingroup hostility effect is based on familiarity, and may not generalise outside of friendship groups over to other social category groups.

Although we found no evidence of an intergroup hostility bias in regard to physical escalatory behaviours, we did find an intergroup hostility bias for non-physical escalatory behaviours. Over one quarter of all intergroup actions were non-physical escalation in comparison to less than one-in-ten intragroup actions. Before proceeding, it is important to first dissect how non-physical escalation may function in antagonistic encounters, which in turn may explain differences at the group level. In the non-verbal communication literature, non-physical presentations of aggression—for example, gesturing, posturing and threats—are not purposeless actions. Rather, these are important communicative ‘displays’ that can indicate status, enforce dominance and thus negate the recourse to violence (Burgoon & Dunbar, 2006; Kemeny, Gruenewald, & Dickerson, 2004; MacLean, 1990; Martens, Tracy, & Shariff, 2012). In this line of research, if the superior status and fighting ability of an opponent are unambiguous, non-physical displays of dominance tend to be short or unnecessary (Kemeny et al., 2004; Martens et al., 2012). This is because the weaker party concedes or withdraws, and this communication is recognised by the other party (Keltner, Young, & Buswell, 1997). However, when this information is missing or ambiguous, or both sides are equally matched, these forces may become obstinate, resulting in further aggressive

‘displaying’ until one party concedes or a physical fight ensues (Burgoon & Dunbar, 2006; Kemeny et al., 2004; MacLean, 1990; Martens et al., 2012).

Although not formally stating a hypothesised distinction between intragroup or intergroup conflicts, this line of research again indicates that familiarity of the opponent may be of predictive importance. Returning to the NTE, we have already reviewed qualitative evidence that shows that friendship fighters are generally conversant with the potential physical risk of fighting fellow acquaintances. Therefore, we might assume that since these ingroup fighters tend to know each other’s relative fighting ability and group status, there would be less functional necessity of these non-physical escalatory displays in intragroup conflicts—unless both parties are equally matched. By comparison, individuals engaging in intergroup conflicts are presumably less familiar with their opponent and may use non-physical displays of aggression to enforce dominance, to size up their opponent or to deter the other party from further escalation (Burgoon & Dunbar, 2006; Kemeny et al., 2004; Martens et al., 2012). Interestingly, examination of the barycentric coordinates plot shows few cases with large proportions of both physical escalation and non-physical escalation—and these only occur for intergroup interactions. This is potential behavioural evidence that in the case of intragroup fights, opponents can circumvent the non-verbal niceties, and proceed to physical fighting. However, for intergroup fights, a ‘displaying’ ritual has communicative importance, which can either stop aggression short of violence or fail (Burgoon & Dunbar, 2006; Kemeny et al., 2004; MacLean, 1990; Martens et al., 2012). In short, there is presumably less information on status and fighting ability available between NTE groups than within NTE groups. Non-physical displays may function to ascertain this information or to deter an unfamiliar other. Therefore, there is a causal expectation of higher levels of intergroup, as opposed to intragroup, non-physical escalatory displays, as evidenced within this study. However, this hinges on familiarity. Therefore, it would be of interest to also examine levels of intragroup non-physical escalation between intragroup strangers in self-

categorised groups. The non-verbal communication literature above predicts that a reduced familiarity between opponents makes non-physical escalatory displays more functional, and therefore more likely.

In sum, our findings indicate that the concept of an intergroup hostility bias needs to be reconfigured to account for specific escalatory behavioural types. The evidence presented above suggests a greater likelihood of intergroup gesturing, posturing and aggressive displays than intragroup examples of the same behaviour. This bias may occur due to an unfamiliarity with the outgroup opponent, and the need to assess one another's level of threat and intentions. However, when physical violence does occur, the behavioural evidence shows that ingroup members are just as likely to be physically violent to one another in conflict as towards outgroup members in conflict. This overall similar level of within- and between-group physical violence persists, even though the succeeding intragroup action tends to be de-escalation over further escalation, and the succeeding intergroup action tends to be escalation. Further examinations are required to better disentangle this relationship.

In an extension of Study 1, the current study re-examined the relationship between the number of active bystanders present, and the number of escalatory acts for videos depicting intragroup fights and intergroup fights. We found no relationship between the number of bystanders and the number of physical escalatory acts in intergroup videos. There was a moderate positive relationship between the number of bystanders and the number of non-physical escalatory acts in intergroup videos. Again, although appreciating relationship does not equal causation, these findings indicate that larger audiences in intergroup conflicts are not related to an increased likelihood of physical violence. Rather, our results suggest that audience presence in intergroup fights is only related to a larger number of threats, insults and postural displays. This may still be taken as behavioural evidence for heightened peer group excitation and impression management concerns during intergroup fights (Felson, 1982;

Jankowski, 1991; W. B. Sanders, 1994; Tomsen, 1997; Weenink, 2014) but the heightened aggression is restricted to non-physical displays.

For intragroup videos, there was a moderate negative relationship and a moderate marginally non-significant negative relationship between the number of bystanders and the number of physical escalatory acts and non-physical acts respectively. Although we still cannot make any definitive claims about causation (e.g., fights with less escalation could attract a larger audience), it is plausible that the presence and actions of ingroup members contribute to this relative violence reduction (e.g., through direct interventions, carer roles or through norm violation signalling; Groff, 2014; Levine et al., 2012; Pedersen et al., 2016).

In Study 1, we ascertained that third-parties were more than a generic feature of the ambient context, but instead contributed 2357 (66.3 %) of the total 3555 actions occurring within conflicts. In this study, we further examined the standardised number of actions contributed by each role: protagonists, targets and bystanders. We expected that the average protagonist and target would perform significantly more actions than the average bystander. We found evidence for the former (H3a), but not the latter (H3b). Protagonists did contribute significantly more actions on average than bystanders, but targets did not. In line with our final prediction (H4), we showed that protagonists contributed significantly more actions than targets.

The finding that protagonists perform more actions than both targets and bystanders might suggest that protagonists are the players most central to a conflict, as is in line with most models of aggression and violence that tend to be offender focused (Anderson, 1997; Finkel, 2014; Slotter & Finkel, 2011). However, in the context of public space violence, it is misleading to place all the emphasis on protagonists alone. Targets and bystanders are active contributors (both roles have an average standardised z score close to zero) who are still found to contribute largely to the stream of actions.

It is interesting to consider why protagonists are most active and why targets and bystanders, on average, contribute similar level of actions. As CCTV examinations of violent human conflict is a recent and developing field (Lindgaard & Copes, 2017), it is difficult to provide any established empirical evidence for these results. Anecdotally, however, from viewing the footage and from experience working on further CCTV projects around bystander intervention, protagonists do indeed often appear to be most active. Protagonists drive the initial confrontation and then either persist in aggressive interactions with the target or interact with those around who are attempting to subdue/incite them. Targets can also be very active during conflicts, from interactions (either de-escalatory or escalatory) with the protagonist and others. However, at times, targets are seen to be side-lined away from danger by concerned bystanders. These bystanders then tend to interact with the protagonist and each other while the target is relatively idle.

The diverse level of bystander involvement found in our study is also quite apparent anecdotally when observing CCTV footage of violence. Concerning inactivity, bystanders are frequently observed to make brief attempts to de-escalate or escalate a conflict and then quickly give up. Alternatively, bystanders are frequently observed to persistently go back and forth between the protagonist, the target and others involved to mediate the situation or to aggressively join in. These diverse levels of involvement are reflected in the data, which show that both the lowest and highest standardised number of actions came from the bystander role. As such, the role of 'active bystander' itself may be oversimplified, ignoring the potential for differing levels of activity and divergent intervention strategies. Therefore, in our final study (Study 7), we further dissect the different bystander roles, and propose a new typology of bystander intervention based on pacifying and policing roles.

5.2 Study 7: Policers and Pacifiers—A New Typology of Third-party Intervention in NTE Conflicts

Conflict is an inevitable by-product of communal-living (von Rohr et al., 2012). When conflict arises, it can destabilise societies and threaten the benefits of cooperative life (de Waal, 2000). For the individuals involved, conflict can damage relationships, inflict trauma and physical harm, and in the most extreme cases death (Flack et al., 2006). However, conflicts in both human (Collins, 2008) and non-human primate societies (Krakauer et al., 2011) seldom result in severe injury. The investigation of how everyday human conflicts are prevented from severe escalation remains largely unexplored. This neglect is a combination of the ethical challenges of experimentally studying human violence (Levine et al., 2011), and of the practical challenges of capturing everyday human conflicts as they occur in real-time (Lindgaard & Copes, 2017). By way of contrast, the greater opportunity to conduct systematic observations in captive and wild animal populations has allowed non-human primate researchers to conclude that severe violence is mitigated by well-developed conflict management mechanisms (Krakauer et al., 2011). One such mechanism is the role of third-parties in creating the conditions for natural conflict resolution. For example, post-conflict, third-party non-human primates affiliate with former quarrellers and engage in consolatory behaviours that reconcile community members and restore the relationships damaged by aggression (Romero, forthcoming). Third-parties also enter ongoing conflicts and interrupt the aggression between opponents (Flack et al., 2005b). These interventions are more prevalent in times of social instability suggesting that the primary function of conflict intervention is to increase social stability, consistent with the wider notion of ‘community concern’ (von Rohr et al., 2012).

However, the dominant discourse in traditional social psychology (and some sociology literature) tends to view the presence of others as a negative (rather than a stabilising) effect on aggressive behaviour and conflict severity. For example, research on the bystander effect

(Fischer et al., 2011; Latané & Darley, 1970) shows that increases in third-party presence results in a greater diffusion of responsibility, and a decreased likelihood of prosocial intervention in emergencies. Interviews conducted with violent offenders conclude that the co-presence of an audience sets a stage in which conflict between opponents is a public spectacle (Weenink, 2014). Peer group excitation and intense feelings of group membership can result in ‘frenzied attacks’ (Weenink, 2014), ‘forward panics’ (Collins, 2008) and ‘wolf-packing’ (Keiser, 1969), all phenomena in which a weaker victim is severely attacked with uncontrollable and excessive force (Jankowski, 1991). Felson (1982), from interviews with former psychiatric patients and prisoners, concluded that the mere presence of others, and the impression management concerns they invoke in the individual, is sufficient to increase the likelihood of severe violence. In common, all of these perspectives propose that the co-presence of third-parties is critical *for conflict escalation* and *not* for natural conflict resolution.

The limitations of interview studies, however, are well known. Participant response bias implies that more desirable answers are expressed, while retrospective recalling of violent episodes may be inaccurate owing to cognitive constraints, self-deception and memory failure (Laney & Takarangi, 2013; Saunders, 1991). Offender interviews are also restricted in that they tend to represent cases in which a person was arrested or prosecuted. These incidents, therefore, are likely to be high in severity and are less likely to depict resolved conflicts involving third-parties, which would remain largely unreported. As such, these recalled events may over-represent the negative aspects of third-parties (coaxing, partisan support, etc.) and not capture the potential existence of third-party natural conflict resolution found so often in the non-discriminative observations of non-human primate populations.

To circumvent the limitations associated with interview accounts, and the practical challenges of capturing everyday human conflicts as they occur in real-time, Levine and colleagues (2011) and more recently the thesis author (Studies 1 and 6), conducted systematic,

behavioural analyses of the actions of individuals captured in CCTV footage of street fights. This CCTV footage—comparable to observations in non-human primate research—provided unobtrusive and naturalistic access to the behaviours of individuals in and around real-life aggressive episodes. This dataset (although still biased towards disturbances which gained the camera operatives' attention) also captured conflicts of varying severity; from mild altercations and posturing that dissipates naturally, to incidents that escalate to more severe outcomes.

To investigate the role of third-parties in shaping conflict outcomes, Levine and colleagues (2011) and the thesis author (Study 1) coded the escalatory and de-escalatory acts of nearby bystanders who contributed to a public-space conflict. Contrary to the idea that third-parties are detrimental to the violent outcome, analysis of the observable behaviours showed a third-party tendency to contribute de-escalation over escalation. Furthermore, contrasting with established research on the bystander effect, increased numbers resulted in a greater likelihood of third-party intervention and a higher number of de-escalatory attempts. In the case of Levine et al. (2011), it was these accumulated de-escalatory attempts provided by multiple bystanders which best predicted reduced (as opposed to increased) conflict severity. The counter argument to emerge from these behavioural analyses is that the co-presence of third-parties is critical *for natural conflict resolution* (as advocated in the non-human primate literature) and *not* conflict escalation (as proposed by offender interview work).

While the natural human conflict work of Levine and colleagues (2011) is novel in both empirical findings and methodological approach, it failed to account for the specialised roles third-party individuals adopt when intervening in conflict. Rather, the term third-party referred to any active bystander (not of the original aggressive dyad) who performed at least one escalatory or de-escalatory action towards another. Also, whilst Study 6 noted a large discrepancy in bystander levels of involvement, no distinction was made between bystanders

who interacted with the aggressive dyad, and those who did not. In the non-human primate literature, third-party intervention is a division of labour with degrees of specialisation that have a considerable impact on conflict dynamics and the severity outcome (Krakauer et al., 2011). For example, Flack and colleagues investigating the third-party conflict management strategies of pigtailed macaques (Flack et al., 2005a, 2005b, 2006) differentiated between ‘pacifier’ and ‘policer’ roles (Krakauer et al., 2011). A pacifier was defined as a third-party bystander who intervened towards one of the conflict opponents, but not both. A policer was defined as a third-party bystander who targeted both conflict parties indiscriminately (Krakauer et al., 2011). Comparing third-party pacifier and policer interventions, Flack and colleagues (2005a, 2006) showed that policing is a more effective strategy than pacifying at reducing the intensity of aggressive conflicts. Policing, and not pacifying, was also capable of terminating a conflict entirely – but only when performed by the individuals recognised as most powerful (Flack et al., 2005a, 2005b). Using an epidemiology analogy and computational modelling, Krakauer et al. (2011) further argued that policing, and not pacifying, could provide ‘social conflict immunity’. Policing contained conflict contagion by allowing the minimal number of interveners to dissipate the maximum number of aggressors. When policers policed, the intensity of conflicts tended to abate. When policers were experimentally removed (both in captive populations and computationally), there was an increase in the number of conflicts and the intensities of these conflicts.

In a similar vein, von Rohr et al. (2012)—in their work on third-party interventions across four captive chimpanzee populations—also proposed a typology of intervention: ‘partial bystander involvement’ and ‘impartial bystander intervention’. In this distinction, a partial bystander provided partisan support to one of the two contestants. An impartial bystander, (also referred to as an arbitrator), however, intervened indiscriminately to separate both opponents of the ongoing contest. In this dataset, impartial arbitrators, like policers, were found to be critical players at stifling conflict escalation. From 69 impartial interventions,

arbitrators were found to terminate conflicts in 89.96 % of occasions (von Rohr et al., 2012). Furthermore, this success rate improved to 100 % when more than one arbitrator simultaneously contributed to the impartial intervention. Effective impartial interventions have also been recorded in Bornean orangutans (Tajima & Kurotori, 2010), mountain gorillas (Sicotte, 1995) and in additional populations of rhesus macaques (Beisner & McCowan, 2013). Beisner and McCowan (2013) concluded that impartial interventions aimed at both parties were positively associated with conflict termination, and negatively associated with outcome severity and group-level trauma. In sum, the typology of conflict roles in the non-human primate literature extends beyond the dichotomy of ‘opponents’ and ‘third-party bystanders’. Rather, third-party intervention is a division of labour with degrees of specialisation that have considerable consequences on the conflict dynamics. It is the specialised role of a third-party bystander, and not the presence of a third-party bystander, which is best predictive of conflict severity.

In everyday human conflicts, the equivalent classification of pacifier (henceforth, an intervener who attempts to de-escalate *one* of *either* the protagonist or the target, but not both) and policer (henceforth, an intervener who attempts to de-escalate both the protagonist *and* target) is absent. However, a common term given to those who attempt to arbitrate human conflict is ‘mediator’. Mediators are, for the most part, associated with a reduction in conflict escalation (Moore, 2014). For example, while Felson (1982) largely assumed that mere presence of others increased the severity of violent conflicts, he did specify that third-party mediators could reverse this effect, by offering fighters the opportunity to step down while saving face. Research on gang violence stresses the importance of ex-gang members to act as legitimate mediators in ongoing violent conflicts (Slutkin, Ransford, Decker, & Volker, 2015). Further research into peer conflict in educational settings (Schellenberg, Parks-Savage, & Rehfuss, 2007) and corporate and legal disputes (Moore, 2014), also highlights the necessity of mediator presence to reach conflict settlements. However, there is a lack of precision, as

the term ‘mediator’ remains largely unspecified. For example, Felson (1982) allowed interviewees to arrive at their own definition of mediator. It is therefore uncertain whether these mediators addressed one party or both parties, whether they used de-escalatory behaviours or escalatory behaviours. During gang violence interventions “mediations occur through many techniques such as meeting one-on-one with aggrieved individuals, hosting small group peace-keeping session to foster diplomacy” (Slutkin et al., 2015, p. 3). Here a mediator takes on a number of tasks which may encompass both ‘pacifying’ and ‘policing’ roles. Crucially, in the non-human primate literature, it is this distinction of specific mediator roles that has predictive power over the conflict severity outcome. Even if the term mediator is deemed sufficient in the human conflict literature, there still exists no systematic examination of the effectiveness of mediators (both ‘pacifiers’ and ‘policers’) in reaching natural conflict resolution in everyday examples of aggressive conflicts. Our initial goal, therefore, is to determine the potential existence of pacifier and policer third-party roles in everyday human conflicts captured by CCTV. Our overall aim is to conduct the first systematic study that assesses the effectiveness of these conflict management strategies in reducing human conflict severity.

5.2.1 Hypotheses

With former evidence that third-parties in public space violence largely attempt to placate conflicts (Levine et al., 2011; Liebst et al., 2017; Parks et al., 2013; Study 1), we predict:

H1: that third-party pacifiers (H1a) and policers (H1b) will be identified in our footage.

Following Levine et al. (2011), we predict:

H2: that the number of active third-party bystanders will be a significant negative predictor of conflict severity.

However, we expect that the number of active bystanders will lose its predictive capacity once accounting for specific intervention roles. Therefore, we predict:

H3: that the negative relationship between the number of active third-party bystanders and conflict severity will be attenuated by the number of pacifiers (H3a), and negated by the number of policers (H3b).

As it has been shown analytically and computationally that policing is a more efficient strategy than pacifying in non-human primates (Krakauer et al., 2011), we predict:

H4: that number of policers will be the strongest predictor of violence severity in our CCTV footage.

5.2.2 Methods

We performed secondary data analyses on the 43 CCTV video clips of public-space conflicts previously coded in Study 1. An active bystander was defined as any individual who was actively involved by either performing or receiving at least one escalatory or de-escalatory behaviour over the course of a video clip. A pacifier was an intervener who attempted to de-escalate *one* of *either* the protagonist or the target, but not both. A policer was an intervener who attempted to de-escalate *both* the protagonist *and* the target. As such, while pacifiers and policers were also considered active bystanders, pacifiers and policers were mutually exclusive.

Outcome severity was operationalised as the severity of violence observed in a clip. This variable had three levels, mild, moderate and severe (see code book, Appendix). Violence was mild if there was some posturing and aggressive gesturing, light pushing and shoving and feinted movements, but no physical punches, kicks, slaps, etc., were landed on an opponent. Violence was considered moderate if a physical fight took place, punches, kicks, slaps, etc., were landed, but nobody was badly hurt, knocked unconscious or kicked

repeatedly. Violence was considered severe if a person was badly hurt, bled profusely, was knocked unconscious or kicked severely while on the floor.

5.2.3 Results

Of the 43 CCTV clips, nine incidents depicted mild violence, 24 incidents depicted moderate violence, and ten incidents depicted severe violence. Across the 43 incidents, there were 43 protagonists who drove the initial aggression, and 43 targets who were the recipients of that initial aggression. Protagonists and targets represented the aggressive dyad. Across the 43 incidents there were 246 active third-party bystanders ($M = 5.72$, $SD = 3.06$), 119 of whom were pacifiers ($M = 2.77$, $SD = 1.72$), and 64 of whom were policers ($M = 1.49$, $SD = 1.39$). As such, pacifiers (H1a) and policers (H1b) were identified in everyday human conflicts, though policers were less frequent.

Table 5.2

Bivariate correlations of bystander role measures and severity of violence

	1	2	3	4
1 Severity of Violence	-			
2 Number of Bystanders	-.425**	-		
3 Number of Pacifiers	-.305*	.759***	-	
4 Number of Policers	-.472**	.482**	.099	-

Note. $N = 43$; * $p < .05$, ** $p < .01$, *** $p < .001$

Table 5.2 shows the relationships among all predictor variables—the number of bystanders, the number of pacifiers and the number of policers—and the outcome variable—the severity of violence. All predictor variables showed an inverse relationship with violence severity, and the number of policers was the strongest predictor of this outcome variable. Possibly indicating a threat of regression model collinearity, the two predictor variables—the number of bystanders and the number of pacifiers—were highly correlated²⁰, Pearson's

²⁰ Field (2009) suggests that where a correlation between any two predictors exceeds .90 (or more conservatively .80), those predictors should not be included in the same regression without good theoretical justification for doing so. Tabachnick and Fidell (2007) set a more stringent criteria of .70. Our correlation value

$r = .759, p < .001$. It is important to stress, however, that highly correlated predictor variables do not necessarily equate to multicollinearity presence in a model. Therefore, we checked whether multicollinearity was present in our models by assessing the tolerance and variance inflation factor (VIF) collinearity statistics. These statistics suggested that the model predictors posed no threat of multicollinearity²¹.

A hierarchical multiple regression with three steps was conducted to estimate the relationship between the number of bystanders present, the number of pacifiers present, and the number of policers present on the severity of violence recorded in an incident (see Table 5.3). Model 1 included the number of active bystanders present in an incident as the single predictor of violence severity. Model 2 retained the number of active bystanders present and further included the number of pacifiers present in an incident as an additional predictor of violence severity. The final model (Model 3) retained the number of active bystanders present, the number of pacifiers present, and further included the number of policers present in an incident as an additional predictor of violence severity.

of .759 is below the criterion thresholds set by Field (2009), but exceeds the more conservative .70 value of Tabachnick and Fidell (2007).

²¹ Menard (1995; as cited in Field, 2009) postulates that tolerance values below .20 flag potential multicollinearity concerns, with values below .10 indicating serious multicollinearity issues. None of the tolerance values in any of our models were below these thresholds (ranging from .26–1.00). We also checked the VIF statistics of our regression models. Predominately, Field (2009) and others (e.g., Hair, 1998; Kutner, Nachtsheim, Neter, & Li, 2005; Myers, 1990) propose that VIF values greater than 10 indicate the presence of multicollinearity. A more conservative VIF criterion of greater than 5 can also be found in the statistical literature (Rogerson, 2001). The VIF values in our models did not exceed 3.89 (ranging from 1.00 to 3.89). This suggests that the two predictors (the number of bystanders and the number of pacifiers) posed little threat of multicollinearity and that the two predictors could be included together in a regression model without concern of biasing our regression models.

Table 5.3

Hierarchical multiple regression analysis of the number of bystander, pacifiers and policers in predicting the severity of violence

Model	Variables	B	SE_B	95% CI	t	p	β	F	R^2	ΔR^2
1	Intercept	1.558	0.201	[1.152, 1.964]	7.751	<.001		9.059	0.181	
	No. of bystanders	-0.094**	0.031	[-0.156, -0.031]	-3.010	.004	-0.425			
2	Intercept	1.553	0.205	[1.138, 1.968]	7.559	<.001		4.442	0.182	0.001
	No. of bystanders	-0.101*	0.048	[-0.198, -0.003]	-2.084	.044	-0.458			
	No. of pacifiers	0.017	0.086	[-0.157, 0.190]	0.193	.848	0.042			
3	Intercept	1.637	0.197	[1.239, 2.035]	8.323	<.001		5.334	0.291	0.109
	No. of bystanders	-0.011	0.058	[-0.129, 0.108]	-0.180	.858	-0.048			
	No. of pacifiers	-0.089	0.092	[-0.274, 0.097]	-0.966	.340	-0.226			
	No. of policers	-0.207*	0.084	[-0.378, -0.036]	-2.451	.019	-0.427			

Note. β = standardized regression coefficient; No. of bystanders = total bystanders; No. of pacifiers = number present; No. of policers = number present;
 * $p < .05$, ** $p < .01$, *** $p < .001$.

Model 1 was statistically significant, $F = 9.06, p = .004$ and explained 18.1 % ($p = .004$) of the variance in violence severity. The number of active bystanders present in an incident was a significant predictor of violence severity. The estimated unstandardized beta (B) denoted an inverse relationship between number of bystanders and the severity of violence. This beta indicated that a one unit increase in the number of active bystanders was associated with a $-.094$ unit decrease in violence severity (where mild = 0, moderate = 1, severe = 2).

Model 2, which included the number of pacifiers and the number of bystanders as predictors of violence severity, was statistically significant, $F = 4.44, p = .018$. This model explained 18.2 % of the variance in the data, which was a marginally small, non-significant improvement ($+0.1\%$, $F \text{ change} = 0.037, p = .848, n.s.$) in predictive capacity from Model 1. In Model 2, the number of active bystanders remained a significant negative predictor of the outcome variable violence severity, when controlling for the number of pacifiers present. The estimated unstandardized beta indicated that when controlling for the number of pacifiers, a one unit increase in the number of active bystanders was associated with a $-.101$ unit decrease in violence severity. The number of pacifiers present in an incident had no relationship to violence severity.

The final model (Model 3) assessed the relationship between all predictors, the number of active bystanders, number of pacifiers, and the number of policers present, on violence severity. This model was also statistically significant, $F = 5.334, p = .004$. The model accounted for 29.1 % of the variance in the data, which was a 10.9 % improvement ($F \text{ change} = 6.01, p = .019$) in predictive power from Model 2. As such, the number of policers present in an incident significantly contributed to the final model. The number of active bystanders present was no longer a significant predictor of violence severity when controlling for the number of pacifiers and the number of policers present. The number of pacifiers present remained a non-significant predictor of violence severity in this final model.

The number of policers present was a significant predictor of violence severity when controlling for the number of bystanders and the number of pacifiers present. The estimated unstandardized beta indicated, that when controlling for the number of active bystanders and pacifiers, a one unit increase in the number of policers was associated with a $-.207$ unit decrease in violence severity.

5.2.4 Discussion

Everyday conflict is an unavoidable feature of societal living. Although aggressive encounters can result in severe violence and injury, everyday conflict for both human and non-human primates rarely escalate to such devastating degrees. Research investigating conflict management mechanisms in non-human primates demonstrates that third-parties have specialised intervention roles that are important for achieving natural conflict resolution. Specifically, when conflict breaks out between dyadic opponents, a third-party ‘policer’ can provide impartial interventions which stifle conflict escalation and reduce the potential for severe harm. Pacifiers who target one of the aggressive dyad, and individuals who do not address either of the aggressive dyad, are significantly less effective at alleviating conflict (Krakauer et al., 2011). Yet, owing to the challenges of systematically studying everyday human conflict, little is known about how *human* conflicts may be managed by third-parties. The aim of this study was to conduct a CCTV behavioural analysis of bystander interventions in public-space conflicts, to identify the potential existence and effectiveness of third-party specialised intervention roles in human societies.

We demonstrated that active third-parties, pacifiers (H1a) and policers (H1b) are present in everyday human conflict. Contrary to the perspective that the presence and actions of third-parties is likely to exacerbate conflict and increase severe frenzied violence (Weenink, 2014), our results showed that the number of active bystanders was negatively associated with violence severity (H2). This is in line with the previous findings of Levine and colleagues (2011), from which we share a 21 % CCTV sample overlap. Contrary to our

predictions, we did not find that the negative relationship between the number of active third-party bystanders and conflict severity was attenuated by the number of pacifiers (H3a).

Rather, the number of pacifiers was not predictive of violence severity when controlling for the number of bystanders (in Model 2), and when further controlling for the number of policers (in Model 3). We did, however, demonstrate that the relationship between the number of active bystanders and violence severity was negated once policer roles (H3b) were accounted for. A final predictive model, which included the number of active bystanders, pacifiers and policers, showed that only an increase in the number of policers was significantly, negatively related to violence severity (H4).

Previous research showed that the greater likelihood of third-parties to carry out de-escalatory (rather than escalatory) behaviours predicted a reduction in violence severity (Levine et al., 2011). Our results confirmed that the de-escalatory actions of third-parties may be important for conflict resolution and not violence escalation—however, only if those third-parties contained policers who actively attempted to de-escalate both members of the aggressive dyad. In our data, pacifiers, who were also defined based on their de-escalatory attempts (albeit to one of the aggressive dyad) were not inversely related to violence severity. One explanation for this null finding could be that pacifiers *are* effective peacekeepers; however, they undertake the unpleasant task of breaking up conflicts that are already severe. Specifically, it may be physically unfeasible to perform a policer role (i.e., de-escalating both parties), once violence has reached severe levels—leaving pacifying as the only available strategy.

Another explanation for this null finding is that by attempting to de-escalate just one party, a pacifier may be perceived as a ‘partial’ supporter of one opponent, as opposed to an ‘impartial’ bystander who is acting indiscriminately (von Rohr et al., 2012). Tentatively, given that a pacifier may be perceived as acting on one’s opponent’s behalf, their placatory signals may be ignored by the non-de-escalated other. By way of contrast, by acting indiscriminately,

policers may be perceived as impartial, and therefore their de-escalatory attempts may face less opposition (Krakauer et al., 2011; von Rohr et al., 2012). Here, impartial de-escalatory interventions may send explicit signals to fighters that their aggressive behaviours, as perceived by a (potentially neutral) outside party, are unacceptable (Groff, 2014; Reynald, 2017).

An integral piece of information not captured in our CCTV data is the power relationships that exist between individuals. In the non-human primate literature, policing is found to have varying degrees of success depending on the hierarchical structure of the society; with asymmetrical alpha dominated societies having greater policing success rates than more democratic animal groups (where there is less agreement regarding the main sources of authority) (Flack et al., 2005b). As such, policing tends to be performed by the few, and these few tend to be high-power individuals (Flack et al., 2005b; von Rohr et al., 2012). Recognised power provides both legitimacy to control the conflict, and license to intervene at a low personal cost to the policer (von Rohr et al., 2012). By way of contrast, interventions by low-power individuals, or individuals in democratic societies, tend to be less effective and potentially more costly to the intervener (Flack et al., 2005b; von Rohr et al., 2012).

In our CCTV footage, it could be that those who employed successful policing strategies also tended to be those with sufficient recognised power and perceived legitimacy to intervene (Haslam, 2004; von Rohr et al., 2012). It is further possible that human interveners who are not recognised as powerful tend not perform a policing intervention, or may run a higher risk of intervention failure or of victimisation if they do (Flack et al., 2005b; von Rohr et al., 2012). Although here we make gross associations between human and non-human primate social relationships; comparative research does demonstrate how vertical social hierarchies in humans, mirroring those found in non-human primates, can explain similar levels of cooperation between species (Cronin, Acheson, Hernández, & Sánchez, 2015). Comparative research further shows how social closeness between human individuals,

and non-human primate individuals alike, can predict similar patterns of post-aggression consolation (Lindgaard et al., 2017). Therefore, it is important to entertain the possibility that, in human contexts, as in non-human primate contexts, the effectiveness and costs of policing may be mediated by social perceptions of the policer.

Whether policing interventions produce conflict alleviation may also be qualified by whether the aggression happens in intragroup or intergroup contexts. In intragroup contexts of public space violence, opponents and interveners tend to belong to same friendship or acquaintance group (Levine et al., 2012; Liebst et al., 2017). Here, the fighting opponents may be familiar with a policer's general intentions and relative social group power (see discussion on familiarity, Study 6). With less ambiguity, we might predict that policing interventions in intragroup contexts tend to be successful. As it is important to preserve group stability and to maintain relationships among familiar intragroup individuals (de Waal, 2000) we might also predict that policing in intragroup contexts is relatively common. During intergroup conflicts, it is plausible that third-party policers from an opposition group may be perceived as partisan supporters rather than impartial interveners, or perceived as legitimate targets of aggression interchangeable with the outgroup opponent (Liebst et al., 2017). As such, policing interventions in intergroup contexts may be less successful with a higher cost attached. Finally, policing interventions for both intragroup and intergroup contests may also come from external third-parties, entirely unrelated to any of the conflict participants. Although is difficult to predict the effectiveness of external third-party policing, presumably the perception of impartiality should be high, and the likelihood of opponent entitativity low. Unfortunately, our limited sample size did not allow us to examine the appraisal of social groups as moderators of policing effectiveness and risk.

It must be stressed however, that for all findings, it is difficult to ascertain causation. While it is plausible that the actions of policers effectively stifle conflict escalation (Krakauer et al., 2011), it may also be true that less severe incidents attract more policers—i.e.,

individuals who feel it is safer to intervene and address both parties. This attraction to low-severity incidents explanation, however, runs counter to the aversive-arousal reduction hypothesis (Dovidio, Piliavin, Gaertner, Schroeder, & Clark, 1991), which postulates that the unpleasant empathetic arousal from perceiving another in severe distress is more likely to motivate intervention. This attraction to low-severity incidents explanation also diverges from more recent behavioural findings in the bystander literature which shows that individuals are *as* likely (Fischer et al., 2011) or *more* likely (Lindegaard et al., 2017) to intervene in high as opposed to low danger emergencies. To better disentangle cause and effect, future research could examine at which time point in a conflict police intervene. However, even this may not be entirely conclusive. For example, if police intervene early then it may be taken as evidence that policing interventions stifle conflict escalation. If police intervene at a later time point (presumably because the incident is perceived as not severe), this may be taken as evidence that police intervene due to the inherent low risk. However, in this scenario, it could still be that police at this later time point defused the conflict, and that the incident could have endured, or have escalated into high severity violence had the policing intervention not taken place at that time. Therefore, currently, we can only describe a relationship between police presence and negative violence severity, and cannot ascertain direct causation. Furthermore, we currently do not have any information on what drives an individual to intervene in a police capacity—whether this is an individual difference, a dispositional role, or a behaviour promoted by social-identity processes. Further investigation into these avenues is required before a full understanding of the motivations and effectiveness of specific third-party roles in reducing everyday human conflict can be established.

6 General Discussion

Violence in public spaces remains a persistent societal issue that is widely recognised but inadequately understood. To orientate this problem within the current state of violence literature, this thesis began by evaluating the key metaphors used in psychology, sociology and primatology to explain the emergence of violence. We argued that each literature orientates to the problem through a different organising metaphor. In psychology, violence is seen through the lens of an intrapsychic balance metaphor (Finkel, 2007, 2014; Finkel et al., 2008; Finkel & Eckhardt, 2013; Slotter & Finkel, 2011). Violence occurs when an individual fails to inhibit their own aggressive urges and there is a loss of balance between resources that support control and resources that undermine it. In sociology, the transition to violence is seen through the lens of an interactional transfer of energy metaphor (Collins, 2008, 2009). This theory focuses on the interaction, rather than the individual, and suggests that violence occurs through the overflow or transition of emotional energy around inhibitory impulses that usually keep violence at bay. Finally, the non-human primate literature sees violence through the lens of an economic metaphor. Drawing on cost–benefit analyses, it suggests that violence emerges when the benefits that might promote violence exceed the costs (augmented with a social relationship component) of this potentially risky activity (de Waal, 2000; Georgiev et al., 2013).

We suggested that while these metaphors have some utility at their different levels, they are not metaphors that can easily accommodate the importance of wider interactions that extend beyond the aggressive dyad. We argued that it is necessary to go beyond dyadic interactions because most violence occurs in the presence of third-parties (Felson, 1982; Planty, 2002). These third-parties are more than a generic feature of the ambient context in which violence emerges from an aggressive dyad. Rather, third-parties are participants in the stream of aggressive interactions that precede and succeed a violent outbreak (Levine et al., 2011). Relevant social categories that provide expectations and give meaning to uncertain

events are woven within and between all relevant actors (Otten, 2009; Turner et al., 1987). When group identities are salient and violence occurs, it is shaped by situationally-relevant social identities that govern which behaviours are acceptable and which targets are legitimate (Drury & Reicher, 1999; Reicher, 1984). Further, individual group members tend to support their own group members over others (Reicher & Haslam, 2009; Stürmer & Snyder, 2009). Groups also ‘self-police’ their deviant group members, whose violent actions go against the norms and values of the group (Reicher, 1987; Stott et al., 2007, 2001). Therefore, the presence and actions of third-parties, and shared identity with others, *should* be important for understanding the transition to violence in public space, as well as other, contexts.

6.1 Purpose of the Thesis

The purpose of this thesis was twofold. First, we explored how groups and third-parties may shape the trajectory of violence beyond the aggressive dyad. Second, we investigated the extent to which social categorisation (i.e., intragroup or intergroup dynamics) impacts upon escalatory and de-escalatory behaviours and conflict outcomes. In short, to explain violence escalation and inhibition, we posed the question: do individuals beyond the aggressive dyad and social group categorisations really matter?

6.2 Brief Summary of Results

Seven studies were conducted to ascertain the impact of third-parties, social context and social categories on the behaviours and outcomes of public space violence.

In Study 1 (Chapter 2), we conducted a behavioural microanalysis of a corpus of public space CCTV footage (clips $N = 43$; participants $N = 332$) that captured real-life intragroup and intergroup aggressive events. We coded the escalatory and de-escalatory behaviours of individuals in conflict, and demonstrated that collective group self-regulation took place both within and between groups. However, we found a de-escalatory bias towards one’s own group members and an escalatory bias towards outgroup members. We found that

the majority of all actions coded were performed by bystanders and these tended to be de-escalation. The emergence of the importance of third-party behaviours encouraged us to look beyond the aggressive dyad (protagonist and target) to examine how third-parties may affect violent outcomes. Notable effects of intergroup escalation and intragroup de-escalation biases provided further support for the need to examine social group membership as a unit of analysis in models of aggression and violence in public spaces. Subsequently, we tested the principles of the I³ theory (a dyadic model of aggression) against an alternative social model (which allowed normative influence of others) in a dynamic agent-based modelling environment.

Chapter 3 examined the development of this ABM of aggression and violence. To ensure that we could include social groups and social context, we modelled the dynamics of metacontrast group formation as formulated by Salzarulo (2004, 2006). To demonstrate the robustness and validity of our adapted prototypicality function, we conducted replica experiments of Salzarulo's (2004) original work in Study 2. Individual agents with idiosyncratic opinions and no prior information of social categorisations formed shared prototypical positions on a topic. In line with SCT predictions (Oakes, 1996; Turner et al., 1987), dynamic metacontrast comparisons produced social group formation, consensus, polarisation and superordinate recategorisation.

In Study 3 (Chapter 3), we applied the principle of metacontrast to multidimensional social group formation and physical flocking. Agents derived prototypical opinions on colour dimensions from their local context and formed mobile social groups who moved around together in physical space. Analysis of the make-up of these social groups showed that agents were not randomly clustered; rather, those more similar on colour dimensions were more likely to share social groups and be closer together in physical space. As predicted, social group category boundaries were fluid and agents' prototypes shifted with the changes in the comparative context (e.g., Hogg, 2004; Hogg & Terry, 2000). Psychological distance from a

group prototype, rather than physical distance from the centre of the social group, predicted agent migration.

In Study 4 (Chapter 4), we completed the development of our ABM of violence and used it to compare dyadic accounts of aggression and violence with an alternative social model in which individuals derived prototypical positions on violence from the dynamic local social context. We found more spatial clustering of violence and peace, and stronger intragroup consistencies in violence in the social model compared to the dyadic model in which agents determined their aggressive response from their own individual characteristics and inhibitory capacity. As such, individuals in the social model could be inhibited or uninhibited by the local social context. Social groups significantly differed in their levels of violence more in the social model than in the dyadic model.

In Studies 5a and 5b (Chapter 4), we investigated whether a self-categorisation informed model of violence would produce the intergroup hostility biases predicted by self-categorisation theory (e.g., Otten, 2009; Viki & Abrams, 2013) and demonstrated in our first CCTV behavioural analysis in the context of public spaces. We expected that outgroup interactions would be significantly more hostile than ingroup interactions, owing to a greater polarisation of prototypical positions of violence in intergroup interactions pushing more extreme positions on violence. We found evidence of greater polarisation in intergroup interactions compared to intragroup interactions. However, we found no evidence of an emergent intergroup hostility bias, even when adjusting the metacontrast parameters to increase the likelihood of polarisation in intergroup interactions.

In Study 6 (Chapter 5), we de-aggregated the escalatory behaviours in our CCTV footage into physical escalation and non-physical escalation, and re-examined the intergroup hostility bias. In line with our first CCTV study (Study 1), we found an intergroup hostility bias for non-physical escalation. Further, in line with the findings from our later ABM (Studies 5a and 5b), we demonstrated no intergroup hostility bias for physical escalation. We

also examined the standardised number of actions contributed by each role—protagonists, targets and bystanders—to ascertain the relative involvement levels of bystanders. On average, protagonists contributed significantly more actions than targets and bystanders, and targets and bystanders contributed a similar number of actions.

In Study 7 (Chapter 5), we assessed the relationship between specific third-party conflict management strategies (policers and pacifiers) and conflict violence severity. The number of active third-party bystanders was negatively associated with violence severity. However, this relationship was negated once policer roles (i.e., bystanders who de-escalate both antagonists) were accounted for. Number of pacifiers (i.e., bystanders who de-escalate one of the antagonists) was not predictive of violence severity.

6.3 Theoretical and Practical Implications

Our main findings can be summed up in two major contributions:

- i) The thesis provides new insight into intragroup de-escalation and the collective inhibition of violence.
- ii) The thesis delivers improved understanding of third-party bystander conflict management strategies and the relationship to violence severity.

6.3.1 Intragroup De-Escalation and the Collective Inhibition of Violence

From our CCTV behavioural analyses, we demonstrated that dyadic confrontations in the NTE are regulated by the behaviour of bystanders. We found that bystanders were far more likely to attempt to de-escalate than escalate violence, and that groups tried to self-regulate. Specifically, following acts of intragroup aggression, the default response was intragroup de-escalation. Following acts of intergroup aggression in intergroup fights, over one-fifth of the following acts were intragroup de-escalation. Traditional approaches assume that groups are mostly involved in escalation. Here, our novel findings definitively showed that third-parties tend to favour de-escalation in both intragroup and intergroup conflict

situations. Even more encouragingly, this tendency towards nonviolence was recorded in a data sample that is arguably biased towards more serious confrontations (i.e., those that gained the camera operatives' attention).

Not only did we have behavioural demonstrations of self-policing in our CCTV analyses, but using our agent-based model we are beginning to identify some of the potential underlying group-level process mechanisms that may facilitate collective inhibition. Based on our analyses we suggest that these mechanisms can be described as *omniscience of norms* and *the retention of deviants*.

Omniscience of Norms. As a reminder; agents in our simulation did not directly de-escalate one another. Rather, agents determined how they ought to behave by deriving a prototypical position on violence from the agents and social context surrounding them. For this endeavour, agents took into consideration the positions of violence of *everyone* within their local vicinity. With this all-encompassing prototypicality procedure, we found consistent evidence that aggressive agents could be inhibited by their own social groups and local social context—with the proviso that these prototypical positions tend towards peace.

While such omniscience of others' positions on violence may be possible in a simulated world, it is highly unlikely to be true in episodes of real-life public space conflict. Rather, aggressive individuals need to infer the positions of fellow group members, which may be grossly inaccurate. With a lack of information, protagonists may assume that others also support them in their use of violence, a perception that our data largely suggests is untrue (as demonstrated by the majority of de-escalatory behaviour existing in our CCTV data). One might predict that if protagonists had complete knowledge of the general group attitude towards violence (as our simulation agents have), then these individuals could be more effectively inhibited.

In practice, the only way in which an aggressive group member can accurately discern the normative appropriateness of their behaviour—and, by contrast, the only way in which a

group can project what is appropriate to that individual—is through communication and interaction (Postmes et al., 2005; L. G. E. Smith & Postmes, 2011b). A message of nonviolence can be transmitted to a protagonist during a conflict from fellow third-parties—and, in fact, we assume the de-escalatory gestures coded in our CCTV footage are physical representations of this. However, this intervention is *reactive* as opposed to *proactive*. A proactive strategy could be to encourage intragroup dialogue around the legitimacy of violence. Specifically, previous research shows that intragroup dialogue—even in the form of short discussions—can markedly alter group members’ perceptions, stereotypes, attitudes and norms (Haslam, Oakes, Reynolds, & Turner, 1999; Haslam, Turner, Oakes, McGarty, & Reynolds, 1997; L. G. E. Smith & Postmes, 2011b, 2011a). Shared intragroup dialogue allows individuals to actively evaluate their positions on a topic in relation to common beliefs held by the group. This evaluation can result in increased consensus and reduced ambiguity that in turn can alter future behaviours guided by these validations or invalidations (Abrams & Hogg, 1990; L. G. E. Smith & Postmes, 2011b, 2011a). Put simply, open dialogue can establish or re-establish a group norm, and knowledge of this group norm can alter subsequent behaviour (though this is better facilitated through congruent discussions, see L. G. E. Smith & Postmes, 2011b).

Retention of Deviants. A second mechanism that emerges from our work is the importance of the way groups manage the behaviour of deviants. When an ingroup member engages in unfavourable behaviour, the positive social identity for all other group members becomes threatened (Marques et al., 2001). One strategy to alleviate this threat and maintain group cohesion is to ignore or permit the behaviour of this deviant (Otten, 2009). An alternative strategy is to exclude the deviant from the group entirely (Pinto, Marques, Levin, & Abrams, 2010). The results from our CCTV analyses and ABM suggest that group members should adopt neither strategy. More precisely, our CCTV work that examined third-party de-escalatory interventions showed that a failure to address both the protagonist and

target was related to increased violence severity. Results from our ABM (in which group members were grouped together on colour dimensions and never expelled due to deviance) showed that by keeping a more aggressive deviant within the social fold, that deviant could still be inhibited by the social group. The ABM therefore suggests that people should not distance themselves from ingroup deviants or push them out of the group; instead, it is preferable to keep deviants within the group's sphere of influence.

To reactively or proactively align deviant group members towards a more peaceful norm, individuals should always ensure that their shared identity is salient. Proactively, group members should collectively engage in dialogue discussing the content of the shared identity with emphasis on how violence goes against the grain of the identity, which is shown in alternative research (i.e., violence against other nations' football supporters is against the fun-loving Scottish identity [Stott et al., 2001]; or that 'we', as a peaceful protest group, should not engage in violence [Reicher, Stott, Cronin, & Adang, 2004]). Reactively, group members should intervene with de-escalatory behaviours that show social disapproval of violence (Levine et al., 2011; Studies 1, 6 and 7).

Taken together, we argue that it is theoretically plausible that violence reduction can be achieved through increased knowledge of norm disapproval and by retaining deviants within the social sphere of influence. Both assertions can be tested in an agent-based modelling environment. To test these assertions, we could further develop our ABM to treat 'level of omniscience' and 'retention of deviant rates' as adjustable model parameters. By reducing the omniscience parameter, agents will have a less accurate knowledge of other group members' positions on violence. Consequently, we hypothesise that agents' positions on violence will be less in line with the prototypical position of their social groups. This will make it more difficult for agents to be inhibited by their social groups and local social context. We further hypothesise that by reducing the deviant retention parameter, and allowing for individuals to be expelled from their social groups, deviants will be more likely to act in

accordance with their dispositional properties. If these dispositions favour violence, it will likely result in increased violence. Although these predictions are tentative up to this point, they do highlight one of the major advantages of employing agent-based modelling as an empirical tool—the ability to formalise and experimentally adjust model parameters of interest to isolate specific effects and test new hypotheses.

6.3.2 Third-Party Bystander Conflict Management Strategies and Their Relationship to Violence Severity

In 2011, Levine and colleagues proposed that de-escalatory coordination of multiple third-parties best predicted a reduction in conflict severity. However, this work did not specify to whom the de-escalatory behaviours *were* directed, or recommend to whom the de-escalatory behaviours *should be* directed. Further, the authors failed to account for the specialised roles that third-party individuals adopted when intervening in a conflict. From our research that dissected the different third-party roles, we found that the most successful coordination strategy was one that addressed both parties of the aggressive dyad. We found that number of policers (third-party bystanders who de-escalated both parties) was most related—at least in our data—to a less severe outcome. We further found that individuals who did not direct de-escalatory actions towards both of the aggressive parties were not predictive of violence severity. This does not mean that individuals who try to de-escalate one party (pacifiers) are unnecessary for conflict management. For example, it is plausible that pacifiers were effective at conflict resolution, but that they also tended to undertake the unpleasant task of breaking up fights that were already severe. Rather, what we now know, which we did not before, is that having an individual third-party deal with both sides can be important in stifling escalation.

How then can we best cultivate this line of research and inform the development of future bystander intervention initiatives? The simplest take-home message seems to be this:

the best way to intervene in violence is to attempt de-escalation for both protagonist and target. However, this simple proposal should be taken on board with caution.

First, additional research is required to verify whether our findings extend beyond this dataset into further CCTV datasets or alternative conflict domains. Furthermore, no empirical evidence exists that unpacks the mechanisms to explain *why* policing strategies are effective. For example, it could be that policers are effective because the individuals who perform this type of intervention tend to be those in authority or who exemplify the prototypical facets of the group (Flack et al., 2005b; Haslam, 2004; von Rohr et al., 2012). It could also be that timing explains why policing is effective. Specifically, policers may be effective because they intervene while a conflict is in its infancy, when it is more physically feasible to come between two individuals. Therefore, it could be predicted that multiple pacifiers addressing both parties simultaneously (and thus acting as a single policer by proxy) may still be effective at reducing conflict severity if they intervene while the conflict is still embryonic. Further research is required to follow up these lines of enquiry.

Importantly, we currently have no information about the level of risk to policers and pacifiers who attempt to de-escalate conflict opponents. We know from an existing CCTV behavioural analysis by (Liebst et al., 2017) that the risk of victimisation for an intervening bystander (no policer or pacifier role specified) tends to be relatively low. Specifically, in this research, intervening bystanders were victimised on 18 % of occasions, and these victimisations tended to be relatively low in severity (Liebst et al., 2017). As this study relied on police criminal evidence heavily biased towards severe cases, we might expect that the actual risk to policers and pacifiers would be even lower. Further, the level of risk in such events is unlikely to be constant and may increase with shared social group membership between the victim and intervener (Liebst et al., 2017) and perceptions of entitativity with the opponent (Vasquez et al., 2015). Additionally, risk may fluctuate depending on the recognised social power, prototypicality and perceived legitimacy of the intervener (Flack et al., 2005b;

Haslam, 2004; von Rohr et al., 2012). There may also be gender distinctions, whereby it is more socially acceptable (and therefore less provocative) for females to intervene with de-escalatory touching than males (Suvilehto, Glerean, Dunbar, Hari, & Nummenmaa, 2015). Full investigation into these avenues is required before the effectiveness of specific roles undertaken by third-party interveners and the risks attached to intervention can be fully understood.

6.3.3 Policy and Policing Implications

Our central finding that bystanders largely tried to de-escalate conflict and groups largely attempted to self-regulate could have important implications for NTE violence management practices. In times of economic recession, police and public health services find it increasingly difficult to find the resources needed to investigate, prevent and mitigate crime in public spaces (Tipping & Lloyd, 2015). In response, government efforts to reduce antisocial behaviour (including violence) in the NTE have focused largely on reducing and regulating the excessive alcohol consumption of patrons. Despite a raft of government initiatives tailored towards addressing these concerns—including the refereeing of alcohol promotions, implementation of higher alcohol pricing schemes and restriction of alcohol sales to those intoxicated—the evidence regarding the effectiveness of such policies in reducing alcohol consumption or alleviating these antisocial issues is mixed (Hughes et al., 2014; R. Nakamura et al., 2014).

The evidence in this thesis suggests that legislators need to do more than depend on interventions tailored to subdue the interaction between drinking and violence in order to tackle public space violence effectively. Rather, initiatives need to persuade the police and policy-makers of the importance of situational factors in enactments of public space violence.

As has already been established, groups in public spaces tend to be viewed with suspicion and authorities often disperse groups or move them on as a result (Porter, 2015). However, the empirical research taken from real CCTV footage of violence demonstrates that

self-policing can be achieved. It is therefore important to harness the positive power of groups to regulate themselves in context. As a result, bystanders and third-parties may be deployed as an important resource for developing violence reduction initiatives—a measure that could help overextended public resources in times of austerity.

To achieve this feat, the police need to recognise that the default behaviour of groups in the NTE tends towards self-policing those who are behaving in a violent or antisocial fashion. As a result, groups may be willing to disclose information about a deviant member of their group; however, this is unlikely to occur with more assertive and confrontational policing styles. One way in which this opposition can be eased is through a better understanding of identity threat (Stott et al., 2008). In other words, the police need to ensure that attitudes, communications and behaviours towards public groups are sensitive to the identity concerns of these groups. If the police approach the public defensively (as though they are a uniform threat) or characterise them negatively (as lacking agency or rationality), a defensive response will likely be triggered. This defensiveness is due to the perceived identity threat of being misrecognised or unfairly stereotyped by the police and is unlikely to result in cooperation. The ground-breaking work by Stott and colleagues (2013; 2016) showed that open dialogue and displays of mutual respect between authority figures and groups (e.g., football fans or political protest organisations) can be successful in promoting common ground and reducing violence and antisocial behaviour.

Additional CCTV analyses of interactions in NTE settings will further aid evidence-based initiatives and the development of sophisticated policing strategies tailored to the empowerment of public networks. The ultimate gain is increased public safety and the alleviation and reallocation of limited police resources.

6.3.4 Final Evaluation of the Agent-Based Model

When first reviewing the agent-based modelling literature, we were concerned with the criticism that ABMs may simply reproduce the sum of their parts (Leombruni &

Richiardi, 2005). This apprehension was unwarranted as we found that the modelling process provided a great number of generative contributions. At the most basic level, agent-based modelling allowed us to test the principles of an established model of aggression in an iterative fashion. We were able to dynamically compare this established model with an alternative model that included social considerations. The process of formalising the I³ theory (Finkel, 2007, 2014; Slotter & Finkel, 2011) into an agent-based modelling environment allowed us to uncover a theoretical inconsistency that may have been overlooked by the traditional verbal model (e.g., Galán et al., 2009)—namely that impellance and inhibition may not operate orthogonally; instead, the presence of a factor on one side of the equation may represent the absence of a factor on the other side, and vice versa.

Once we tested our social model alternative, we found evidence of emergent collective inhibition and no intergroup hostility bias. Evaluating collective inhibition allowed us to consider ‘omniscience of norms’ and ‘retention of deviants’ as explanatory parameters for future research (see Section 6.3.1, ‘Intragroup De-Escalation and the Role of Collective Inhibition’). A lack of evidence of an intergroup hostility bias from our ABM—which runs counter to our initial predictions and the social identity literature (Otten, 2009; Viki & Abrams, 2013)—was later verified in our secondary CCTV data analysis, where we found no difference in the levels of physical escalation between intragroup and intergroup interactions. This was an unanticipated finding that demonstrates how ABMs are not simply a tool for reproducing what is expected.

To add a social dimension to our ABM of violence, we needed to first model social group formation. To do this, we programmed individual agents to determine social prototypes through metacontrast calculations, with no full expectation of how this command would play out dynamically across numerous iterations. Our model created emergent social groups who flocked together in physical space, even with fluctuations in local neighbourhood compositions and comparative contexts. To our knowledge, this is the first metacontrast-

driven flocking method available. Furthermore, this new method can be easily applied by other simulation researchers who would like to create context-dependent social groups for their own modelling purposes in alternative research domains.

We also documented many informative emergent phenomena of interest that are relevant outside the domain of violence. Studies in static laboratory settings have shown that individuals move closer to others perceived as being ingroup and move further from those perceived as outgroup (Novelli et al., 2010). We modelled these principles longitudinally and dynamically and found that microsteps towards those perceived as contextually prototypical can create segregation away from ‘others’ categorised as non-prototypical. This computational work suggests that metacontrast is a mechanism of interest for researchers of microcontact theory, who have assessed how attraction to similar others at the micro level can lead to self-segregation or resegregation on the macro level (Buttny, 1999; Clack et al., 2005; Dixon & Durrheim, 2003; Schrieff et al., 2005).

Another interesting finding from our group formation modelling work was that most agents in our simulations tended not to migrate and remained loyal to their social groups over time. As our agents were not given any prior identification information, or any sense of belonging, attachment or commitment (i.e., motivations) to stay with a group (Ellemers et al., 2002), these findings suggest that identification is emergent when metacontrast is at play. This supplements the social identity literature on identity formation which shows that identification, and indeed social identity, appears to emerge from joint co-action (see Jans, Leach, Garcia and Postmes, 2015; and Koudenburg et al., 2015 for related ideas). Specifically, this dynamic model proposes that motivations to remain with a group are less important, as long as the individual is satisfied with their local prototype. This is another example of a generative contribution that once again illustrates how agent-based modelling can provide outcomes greater than the sum of the parts.

Aspects of the modelling process were, however, challenging. We originally constructed our ABM without thinking about its direct relation to the CCTV footage. This was because we wanted to compare two *theoretical* models—an established dyadic model of aggression and violence against an alternative model with social considerations. However, by building our ABM based on theoretical principals, we found it difficult to directly integrate our ABM work with the CCTV footage. For example, the ABM did not distinguish between the different types of escalation as the footage did. Additionally, the ABM could not distinguish the level of severity, which was an important control and outcome variable in the CCTV analyses. Further, de-escalation in the ABM was the result of a calculation including all those within the local vicinity, while in the footage de-escalation was a behavioural count of pacifying actions. As aforementioned (see discussion of Chapter 4), the theoretical ABM acted as a *cognitive* model. The CCTV footage did not allow us to examine cognitive processes; instead, we inferred what was normative from the behavioural distinctions observed. To make like-for-like behavioural predictions of public space violence that is calibrated against the CCTV data available, an integrated *behavioural model* may be necessary.

Given that a metacontrast orientated model of group violence did not largely correspond to the intergroup behaviours observed in the CCTV footage, it is important to emphasise how the model can be developed for future generations. We are in the fortunate position of having extracted a great deal of information from our CCTV analyses that can be used to furnish a behavioural ABM. For example, we recorded the average number of bystanders in a clip, average number of actions (with a breakdown of these actions), relative composition of social groups, relative contributions of individuals, preceding and succeeding sequence of behaviours, specific roles that individuals perform and the relation of these variables to the severity outcome. This extracted information can be incorporated into an ABM environment as model parameters. Sensitivity analyses can then be performed in which

one, or more, possible combinations of values for each parameter are adjusted, and the impact measured (Railsback & Grimm, 2011; Saltelli et al., 2008; Saltelli, Tarantola, Campolongo, & Ratto, 2004). For example, knowing the frequency of bystander interaction is X and bystanders' probability of a certain action is Y , it is possible to investigate what would happen if the frequency of bystander involvement X increased and Y was set to a higher or lower placatory level. One could also investigate what would happen as we increase the number of bystanders (group size), composition of those groups (ratio of ingroup to outgroup members) or roles that individuals take on (bystander, pacifier or policer). Sensitivity analyses of the modelled behavioural data can identify the critical regions of parameter space that are most important to an outcome variable of interest (Saltelli et al., 2008, 2004)—specifically, the combination of parameters best predictive of violence escalation and conflict resolution.

Sensitivity analyses can also determine whether any model produced is robust (i.e., able to withstand adjustments in parameter space while providing similar conclusions) or fragile (i.e., only predictive if set up in a restricted way). Sensitivity analyses can also guide future research avenues by offering insight into which parameters may warrant empirical priority, an advantage especially welcomed when time and resources are limited. Finally, sensitivity analyses on our behavioural data may have important implications for future experimental planning—for example, to plan experimental conditions, to know which confounds to control for and to assess sample size required for experimental work. Through the integration of this newly available information with the modelling environment already constructed, it becomes possible to record the relative impact of the different elements of our dynamic model (including the interactions of the individuals, the interactions of the groups, whether actions are intra or intergroup directed) in the trajectory of violence.

Of course, there are important group processes that are currently absent from our model. One key example is ESIM (Drury & Reicher, 2000; Reicher, 1996), which emphasises social identity as an outcome of, as well as input into, dynamic social interactions.

Specifically, normative behaviour and category boundaries can be shaped through relational intergroup interactions with formal authority groups (such as the police or bouncer security staff). When authority figures are perceived as acting illegitimately by another group (for example, from coercive use of power, through unjust stereotyping and restriction of legitimate group practices) ‘moderate’ group members may recategorise with the more extreme subgroups under a single collective identity with norms that favour intergroup hostility and conflict escalation.

Although ESIM is certainly a relevant dimension, we purposely excluded the actions of, and interactions with, authority figures from our CCTV coding and ABM. This exclusion was deemed necessary because we were interested in how groups may self-police and collectively inhibit anti-social behaviour without the presence and intervention of formal authority figures. Future models could be adapted to incorporate insights from ESIM to further explore how public groups interact (both pro-socially and anti-socially) with authority groups in public space settings. One first step to examine these processes would be to widen the scope of CCTV data inclusion, to include footage that displays interactions between members of the public and figures of authority (such as the police and bouncer security staff). A future modelling step would be to introduce more formal policing agents into the computational environment and then to assess the relative impact that these authority figures may have on violence escalation in a variety of contexts. Specifically, by incorporating ESIM theoretical principles and the relative propensities of actions between authority groups and public groups, future ABM iterations could be extended to model additional types of public space conflict (such as violence escalation at sporting events, protests and riots).

Along with perceived (il)legitimacy, more emphasis in future work should also be placed on the level at which individuals identify and the use of more inclusive superordinate categories. Through invoking a common group membership under a single group representation (for example, the broader category of ‘football fan’ or ‘country’s national’) the

cognitive and motivational processes that generate ingroup bias and promote within group helping and intervention can extend to new category members (Levine et al., 2005; Reicher, Cassidy, Wolpert, Hopkins, & Levine, 2006). Furthermore, supplanting subordinate categories of ‘them’ and ‘us’ with the all-encompassing superordinate category of ‘we’, may help regulate violence by setting countervailing norms of positive interactions (e.g., ‘we’re all just out for a good time’ or ‘we’re all regulars here, we don’t need trouble’).

Currently, in the CCTV footage, it is impossible to investigate the higher superordinate levels at which individuals identify. Computationally, because of the metacontrast principles we adopt, our ABM has already provided examples of superordinate recategorisation. For example, we modelled superordinate identity formation in Study 2, as the physicists and biologists created the more inclusive category ‘hard scientist’ in the presence of social scientists. Furthermore, we found in our flocking work (Study 3) that two separate colour groups could merge to subsume a shared category in the presence of a third, more extreme colour group. Although absent from our current work, it is entirely feasible in future endeavours to isolate these higher-level situations, and then separately analyse how changes in the comparative context may affect norm formation and violence (de)escalation. Therefore, and in hindsight, the ABM of this thesis may have greater benefitted from the integration of intergroup processes (SCT), intergroup dynamics (ESIM) and the interpersonal approaches (I³ theory and GAM). With a great number of interesting avenues to pursue, the agent-based modelling work of this thesis has reached a port of call and not its final terminus.

6.4 Violence as Choreography: Future Trajectories

In the later stages of this thesis, we have been considering alternative literatures which may support the development of a new approach that incorporates the role of third-parties into accounts of violence. We offer this as an alternative to the established approaches. It will be recalled from the introduction that we associated those established approaches with particular metaphors. These were metaphors of balance, energy transfer and economic calculation.

Rather than perceiving violence as a loss, or flow, or calculation, we think it appropriate to offer our own, alternative organising metaphor. We suggest instead, with reference to the empirical findings of this thesis and future avenues of interest, that violence escalation may be better understood as a (collective) *choreography*, operating between multiple agents and local norms in a particular social context.

We are not the first to make an analogy between social norms and choreography. Using correlated equilibrium modelling, Gintis (2010) proposed that norms are best explained through third-party social mechanisms known as ‘choreographers’. Choreographers are correlating devices that send signals to social actors about how they should coordinate their actions to achieve stability. In the simplest of examples, a traffic light system is a choreographer that coordinates drivers and provides guidance on the expected behaviour of each social actor (Gintis, 2010). The choreographer’s role is that of an assurance mechanism (a norm) in an otherwise uncertain event. This assurance is indirect in that social actors are not required to coordinate directly with each other—they follow the signals of the choreographer. Choreographers are not restricted to automatic or formal systems such as traffic light signals; rather, “choreographers can come in a variety of types: deliberate and spontaneous, individual and group, macro and micro” (Thrasher & Vallier, 2015, p. 948). Choreography can also be achieved through “diverse modes of dialogue [that] produce convergence on common principles and rules” (Thrasher & Vallier, 2015, p. 934)—a definition that exemplifies our suggestion of open dialogue among group members to enhance the omniscience of norms. The only major prerequisite of a choreographer is that it signals normative guidance to social actors in cases of potential ambiguity that can be used to stabilise future behaviour through correlated coordination.

In our work, norms tend to favour the prohibition of violence, but agents act in the absence of complete information. As a direct translation of Gintis (2010), third-party bystanders of conflict may act as choreographers, who furnish information on the normative

expectations of the unfolding event. Third-parties signal normative guidance directly through active interventions and behaviours that seek to escalate or de-escalate the situation. Third-parties also signal passively through non-interventions (recognising that inaction is often understood as a form of action by antagonists; Levine, 1999). As the majority of bystanders act to inhibit aggressive encounters, we expect that third-parties act as mediums through which aggressive individuals can coordinate more de-escalatory actions towards conflict resolution.

Experimental and computational research into the dynamics of social coordination suggests that ambiguity over the intentions and motivations of others largely explains when attempts at synchrony fail (Chater, 2017; Nowak, 2017). Presumably, in examples of NTE conflict, opponents may be unsure about the underlying intentions and motivations of their fellow opponent, thus making the coordination of peaceful resolve challenging. Active third-party choreographers can provide assurance by acting as vessels in which conflict opponents coordinate their behaviour indirectly. The coordination literature therefore offers a theoretical explanation for why policers in our work may be related to reduced severity. Specifically, that policers provide normative signalling and assurance through which aggressive social actors can coordinate indirectly with one another. However, a more intricate analysis of the sequence of actions of policers and pacifiers, and examination of the attempts to coordinate the actions of an aggressive dyad, are required to verify if this analogy is directly applicable to third-party bystander intervention.

In the same body of social coordination research, individuals who are out of sync produce errors that can affect the ability of other individuals to coordinate (Nowak, 2017). In the context of our research, these individuals may be third-party bystanders who are sending alternative signals to the majority. For example, in the work of Levine et al. (2011), when the sequence of actions from third-parties was consistent (e.g., de-escalation, de-escalation, de-escalation), the severity outcome was low and highly predictive. However, when the message

was mixed (e.g., de-escalation, de-escalation, escalation), the conflict outcome was less predictive and had a greater chance of being severe. This would suggest that it is important for third-party choreographers to attempt synchrony not only with an aggressive dyad but also with other choreographers who may be disrupting the correlated coordination by ‘dancing to another tune’. Again, while we currently have no definitive evidence that speaks directly to these predictions, these are interesting alternative bodies of literature that have recently stimulated our interest and will be considered in future work.

While choreography is theorised as helping to coordinate normative behaviour between individuals, and thus enhancing stability over time, the signalling system requires that a number of preconditions are satisfied. First, to successfully cue a norm, the choreography must be clearly projected to the social actors and the source should be considered legitimate (Thrasher & Vallier, 2015). After all, individuals cannot dance in time to the choreographer’s tune if that tune is not recognised. Receptiveness to this normative information is unlikely to be equal across all sources—rather individuals may be greater influenced when the message is projected from an ingroup source, and the norm defines the very membership of that group (Turner, 1991; Turner & Oakes, 1986). In line with ESIM principles, the influence of such key norm-setters is not fixed but may shift in response to group interactions and changes in the comparative context (Stott et al., 2008).

Once a normative signal is received, social actors are required to share at least some prior expectations of the content of that signal and have some degree of confidence that other parties may be willing to follow that cue (Paternotte & Grose, 2013; Thrasher & Vallier, 2015). Even then, correlated coordination may still only likely occur when social actors have a self-interest to coordinate, a predisposition of reciprocity and norm obedience (as Gintis assumes on a global level that most individuals do), or ‘other regarding preferences’ (Gintis, 2010). While these prerequisite conditions are obvious restrictions, they may translate into important recommendations for bystander interventions. If verified behaviourally, the above

recommendations would suggest that third-party bystanders intervening in conflicts should use explicit and coherent signals that can be easily interpretable by all parties. It is further recommended that choreographers attempt to coordinate with one another to ensure that they sing from the same sheet, thus reducing noise and error. As choreographers will more likely be followed when they are perceived as legitimate, it is also recommended that those bystanders with the most authority, or group prototypicality, lead the dance (Haslam, 2004). As ‘other regarding preferences’ can motivate individuals to coordinate, it is important to emphasise ties and commitment to the social group, and how coordination towards de-escalation may benefit all closely related members. Significantly, if norms are choreography and norms tend towards nonviolence, then ‘you gotta dance!’ (Murakami, 1994).

6.5 Concluding Summary

Setting aside future ambitions and alternative literatures that have stimulated our interest, it is important to sum up the main take-home messages from this thesis. This thesis attempted to address the relative neglect of third-parties, social groups and social contexts as important explanatory components in the transition to violence in public space conflicts. We showed that third-parties *are* an integral feature of the dynamics of violence, collectively contributing more actions than the aggressive dyad, and equally as important to violence research as the dyad’s ability to inhibit their own violent impulses. We suggested that ‘third-party bystander’ should not be an umbrella term—instead, bystander intervention is a division of labour with degrees of specialisation that have considerable impact on conflict dynamics and the severity outcome.

Third-parties *are* important for understanding conflict escalation and de-escalation, *as are* social group categories. We demonstrated that escalatory and de-escalatory behaviours in public space conflicts are not randomly directed, but are highly socially bound. What is most evident is that groups have a tendency to de-escalate their own, and self-policing and collective inhibition takes place. Therefore, groups retaining and policing their own deviants

and shared salient memberships among members who favour the prohibition of violence are invaluable resources that need to be understood better and harnessed for effective conflict management. Agent-based modelling—although still an embryonic methodological tool in social psychology—has allowed us to dynamically explore and manipulate some of the underlying mechanisms relevant to these intragroup processes (among other things).

This thesis advocates that the field requires not only an alternative socially informed metaphor with which to structure violence research, but also alternative dynamic methods through which the interaction between all relevant actors across the emerging aggressive episode can be evaluated.

References

- Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology, 1*, 195–228.
<https://doi.org/10.1080/14792779108401862>
- Adang, O. M. J. (2016). A method for direct systematic observation of collective violence and public order policing. *Sociological Methods & Research, 0049124116661578*.
<https://doi.org/10.1177/0049124116661578>
- Anderson, C. A. (1997). Effects of violent movies and trait hostility on hostile feelings and aggressive thoughts. *Aggressive Behavior, 23*, 161–178.
[https://doi.org/10.1002/\(SICI\)1098-2337\(1997\)23:3<161::AID-AB2>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1098-2337(1997)23:3<161::AID-AB2>3.0.CO;2-P)
- Anderson, C. A., & Bushman, B. J. (1997). External validity of ‘trivial’ experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19–41.
<https://doi.org/10.1037//1089-2680.1.1.19>
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27–51. <https://doi.org/10.1146/annurev.psych.53.100901.135231>
- Anderson, C. A., & Carnagey, N. L. (2004). Violent evil and the General Aggression Model. In A. G. Miller (Ed.), *The Social Psychology of Good and Evil* (pp. 168–192). New York: Guilford Press.
- Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression. *Personality and Social Psychology Bulletin, 21*, 434–448.
<https://doi.org/10.1177/0146167295215002>
- APA. (2015). Violence. Retrieved 2 February 2017, from <http://www.apa.org/topics/violence>
- Archer, J., & Browne, K. (1989). *Human aggression: Naturalistic approaches* (Vol. xvi). Florence, KY, US: Taylor & Frances/Routledge.

- Aureli, F., Cords, M., & van Schaik, C. P. (2002). Conflict resolution following aggression in gregarious animals: a predictive framework. *Animal Behaviour*, *64*, 325–343.
<https://doi.org/10.1006/anbe.2002.3071>
- Aureli, F., & de Waal, F. B. M. (2000). *Natural conflict resolution*. University of California Press.
- Bandura, A. (1973). *Aggression: A social learning analysis*. Prentice Hall PTR.
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Baron, R. A., & Richardson, D. R. (1994). *Human aggression*. New York: Plenum Press.
- Beisner, B. A., & McCowan, B. (2013). Policing in Nonhuman Primates: Partial Interventions Serve a Prosocial Conflict Management Function in Rhesus Macaques. *PLOS ONE*, *8*(10), e77369. <https://doi.org/10.1371/journal.pone.0077369>
- Bellis, M. A., Leckenby, N., Hughes, K., Luke, C., Wyke, S., & Quigg, Z. (2012). Nighttime assaults: using a national emergency department monitoring system to predict occurrence, target prevention and plan services. *BMC Public Health*, *12*(1), 1.
- Benjamin, A. J. (2008). The general aggression model as a framework for understanding torture and genocide. *National Social Science Journal*, *31*, 1–8.
- Berkowitz, L. (1969). *Roots of aggression* (1st ed.). Atherton Press.
- Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*, 59–73. <https://doi.org/10.1037/0033-2909.106.1.59>
- Berkowitz, L. (1993). *Aggression: Its causes, consequences, and control* (Vol. xxiv). New York, NY, England: McGraw-Hill Book Company.
- Birks, D., & Elffers, H. (2014). Agent-Based Assessments of Criminological Theory. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 19–32). Springer New York. https://doi.org/10.1007/978-1-4614-5690-2_687
- Black, D. (2014). *The social structure of right and wrong*. Academic Press.

- Borum, R., & Pynchon, M. (1999). Assessing threats of targeted group violence: Contributions from social psychology. *Behavioral Sciences & the Law*. Retrieved from http://scholarcommons.usf.edu/mhlp_facpub/543
- Brennan, I. R. (2011). In Vino Silentium? Individual, Situational, and Alcohol-Related Factors in Reporting Violence to the Police. *Violence and Victims, 26*, 191–207. <https://doi.org/10.1891/0886-6708.26.2.191>
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*, 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*, 475–482. <https://doi.org/10.1177/0146167291175001>
- Brewer, M. B. (1993). Social identity, distinctiveness, and in-group homogeneity. *Social Cognition, 11*, 150–164. <https://doi.org/10.1521/soco.1993.11.1.150>
- Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. D. Ashmore, L. Jussim, & D. Wilder (Eds.), *Social Identity, Intergroup Conflict, and Conflict Reduction*. Oxford University Press, Incorporated.
- Brewer, M. B. (2002). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*, 429–444. <https://doi.org/10.1111/0022-4537.00126>
- Bromley, R., Thomas, C., & Millie, A. (2000). Exploring safety concerns in the night-time city: Revitalising the evening economy. *The Town Planning Review, 71*, 71–96.
- Brown, R., & Capozza, D. (2006). *Social Identities: Motivational, Emotional, and Cultural Influences* (1st ed.). Hove; New York: Psychology Press Ltd.

- Burgoon, J. K., & Dunbar, N. E. (2006). Nonverbal expressions of dominance and power in human relationships. In *The SAGE Handbook of Nonverbal Communication* (pp. 279–298). California, USA: SAGE Publications, Inc.
- Burrell, A., & Erol, R. (2009). Tackling violence in the night-time economy on the ground: Putting policy into practice in England and Wales. *Crime Prevention and Community Safety, 11*, 189–203. <https://doi.org/10.1057/cpcs.2009.13>
- Bushman, B. J., & Huesmann, R. L. (2010). Aggression. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology, Volume One*. John Wiley & Sons.
- Buss, A. H. (1961). *The psychology of aggression* (Vol. x). Hoboken, NJ, US: John Wiley & Sons Inc. <https://doi.org/10.1037/11160-000>
- Buttny, R. (1999). Discursive constructions of racial boundaries and self-segregation on campus. *Journal of Language and Social Psychology, 18*, 247–268. <https://doi.org/10.1177/0261927X99018003002>
- Campbell, D. T. (1965). Ethnocentric and other altruistic motives. In D. Levin (Ed.), *Nebraska Symposium on Motivation* (pp. 283–311). Lincoln: University of Nebraska Press.
- Campbell, M. W., & de Waal, F. B. M. (2014). Chimpanzees empathize with group mates and humans, but not with baboons or unfamiliar chimpanzees. *Proceedings of the Royal Society B: Biological Sciences, 281*, 20140013–20140013. <https://doi.org/10.1098/rspb.2014.0013>
- Chater, N. (2017). The logic of joint action: A virtual bargaining approach. Presented at the International Convention of Psychological Science, 2017, Vienna, Austria.
- Clack, B., Dixon, J., & Tredoux, C. (2005). Eating together apart: Patterns of segregation in a multi-ethnic cafeteria. *Journal of Community & Applied Social Psychology, 15*, 1–16. <https://doi.org/10.1002/casp.787>

- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*, 588–608.
<https://doi.org/10.2307/2094589>
- Collins, R. (2008). *Violence: A micro-sociological theory*. Princeton: Princeton University Press.
- Collins, R. (2009). The micro-sociology of violence: The micro-sociology of violence. *The British Journal of Sociology*, *60*, 566–576. <https://doi.org/10.1111/j.1468-4446.2009.01256.x>
- Colman, A. M. (1991). Crowd psychology in South African murder trials. *American Psychologist*, *46*, 1071.
- Copes, H., Hochstetler, A., & Forsyth, C. J. (2013). Peaceful warriors: Codes for violence among adult male bar fighters. *Criminology*, *51*, 761–794.
<https://doi.org/10.1111/1745-9125.12019>
- Coxeter, H. S. M. (1969). *Introduction to geometry*. New York: Wiley.
- Crawford, A., & Flint, J. (2009). Urban safety, anti-social behaviour and the night-time economy. *Criminology & Criminal Justice*, *9*, 403–413.
<https://doi.org/10.1177/1748895809343390>
- Cronin, K. A., Acheson, D. J., Hernández, P., & Sánchez, A. (2015). Hierarchy is detrimental for human cooperation. *Scientific Reports*, *5*. <https://doi.org/10.1038/srep18634>
- Davis, N. Z. (1973). The Rites of Violence: Religious Riot in Sixteenth-Century France. *Past & Present*, (59), 51–91.
- de Waal, F. B. M. (1982). *Chimpanzee politics: Power and sex among apes* (1st ed.). New York: Harper & Row.
- de Waal, F. B. M. (2000). Primates—a natural heritage of conflict resolution. *Science*, *289*, 586–590.

- de Waal, F. B. M., & van Roosmalen, A. (1979). Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology*, *5*, 55–66.
<https://doi.org/10.1007/BF00302695>
- Decker, S. H. (1996). Collective and normative features of gang violence. *Justice Quarterly*, *13*, 243–264.
- DeDeo, S., Krakauer, D. C., & Flack, J. C. (2010). Inductive game theory and the dynamics of animal conflict. *PLoS Comput Biol*, *6*. Retrieved from
<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000782>
- DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence*, *1*, 245–258.
<https://doi.org/10.1037/a0023842>
- Diener, E. (1976). Effects of prior destructive behavior, anonymity, and group presence on deindividuation and aggression. *Journal of Personality and Social Psychology*, *33*, 497–507. <https://doi.org/10.1037/0022-3514.33.5.497>
- Dixon, J., & Durrheim, K. (2003). Contact and the ecology of racial division: Some varieties of informal segregation. *British Journal of Social Psychology*, *42*, 1–23.
<https://doi.org/10.1348/014466603763276090>
- Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. (1939). *Frustrations and aggression*. Institute of Human Relations, Yale University.
- Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup Bias. In *Handbook of Social Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470561119.socpsy002029>
- Dovidio, J. F., Piliavin, J. A., Gaertner, S. L., Schroeder, D. A., & Clark, R. D. (1991). The arousal: Cost-reward model and the process of intervention: A review of the evidence. In *Prosocial behavior* (pp. 86–118). Thousand Oaks, CA, US: Sage Publications, Inc.

- Drury, J., & Reicher, S. D. (1999). The intergroup dynamics of collective empowerment: Substantiating the social identity model of crowd behavior. *Group Processes & Intergroup Relations*, 2, 381–402. <https://doi.org/10.1177/1368430299024005>
- Drury, J., & Reicher, S. D. (2000). Collective action and psychological change: The emergence of new social identities. *British Journal of Social Psychology*, 39, 579–604.
- Ellemers, N., Spears, R. G., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, 53, 161–186.
- Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4, 41–60. [https://doi.org/10.1002/\(sici\)1099-0526\(199905/06\)4:5%3C41::aid-cplx9%3E3.0.co;2-f](https://doi.org/10.1002/(sici)1099-0526(199905/06)4:5%3C41::aid-cplx9%3E3.0.co;2-f)
- Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.
- Fedurek, P., Slocombe, K. E., Hartel, J. A., & Zuberbühler, K. (2015). Chimpanzee lip-smacking facilitates cooperative behaviour. *Scientific Reports*, 5, 13460. <https://doi.org/10.1038/srep13460>
- Felson, R. B. (1978). Aggression as impression management. *Social Psychology*, 41, 204–213.
- Felson, R. B. (1982). Impression management and the escalation of aggression and violence. *Social Psychology Quarterly*, 45, 245–254. <https://doi.org/10.2307/3033920>
- Felson, R. B. (1993). Predatory and dispute-related violence: A social interactionist approach. In R. V. G. Clarke & M. Felson (Eds.), *Routine Activity and Rational Choice*. Transaction Publishers.
- Felson, R. B. (2009). Violence, crime, and violent crime. *International Journal of Conflict and Violence (IJCV)*, 3, 23–39.

- Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior, 17*(3), 220–228. <https://doi.org/10.1016/j.avb.2012.02.007>
- Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *The Journal of Abnormal and Social Psychology, 47*, 382–389. <https://doi.org/10.1037/h0057906>
- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). Sage Publications Ltd.
- Finkel, E. J. (2007). Impelling and inhibiting forces in the perpetration of intimate partner violence. *Review of General Psychology, 11*, 193–207. <https://doi.org/10.1037/1089-2680.11.2.193>
- Finkel, E. J. (2014). The I³ Model: Metatheory, theory, and evidence. In *Advances in Experimental Social Psychology* (Vol. 49, pp. 1–104). Elsevier. Retrieved from <https://doi.org/10.1016/B978-0-12-800052-6.00001-9>
- Finkel, E. J., Bodenhausen, G. V., & Bushman, B. J. (2008). The I³ theory of aggression. *Unpublished Manuscript*.
- Finkel, E. J., DeWall, C. N., Slotter, E. B., McNulty, J. K., Pond, R. S., & Atkins, D. C. (2012). Using I³ theory to clarify when dispositional aggressiveness predicts intimate partner violence perpetration. *Journal of Personality and Social Psychology, 102*, 533–549. <https://doi.org/10.1037/a0025651>
- Finkel, E. J., & Eckhardt, C. I. (2013). Intimate partner violence. In J. A. Simpson & L. Campbell (Eds.), *The Oxford Handbook of Close Relationships*. OUP USA.
- Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., ... Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin, 137*, 517–537. <https://doi.org/10.1037/a0023304>

- Flack, J. C., de Waal, F. B. M., & Krakauer, D. C. (2005a). Robustness mechanisms in primate societies: A perturbation study. *Proceedings of the Royal Society Biological Sciences*, *272*, 1091–1099. <https://doi.org/10.1098/rspb.2004.3019>
- Flack, J. C., de Waal, F. B. M., & Krakauer, D. C. (2005b). Social structure, robustness, and policing cost in a cognitively sophisticated species. *The American Naturalist*, *165*, 126–139. <https://doi.org/10.1086/429277>
- Flack, J. C., Girvan, M., de Waal, F. B. M., & Krakauer, D. C. (2006). Policing stabilizes construction of social niches in primates. *Nature*, *439*, 426–429. <https://doi.org/10.1038/nature04326>
- Fligner, M. A., & Killeen, T. J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, *71*, 210–213. <https://doi.org/10.1080/01621459.1976.10481517>
- Forsyth, A. J. M., & Lennox, J. C. (2010). Gender differences in the choreography of alcohol-related violence: An observational study of aggression within licensed premises. *Journal of Substance Use*, *15*, 75–88. <https://doi.org/10.3109/14659890902966497>
- Fox, R. (1978). *The Tory islanders: A people of the Celtic fringe*. Cambridge, UK: Cambridge University Press.
- Galán, J. M., Izquierdo, L. R., Izquierdo, S. S., Santos, J. I., del Olmo, R., López-Paredes, A., & Edmonds, B. (2009). Errors and artefacts in agent-based modelling.
- Ge, W., Collins, R. T., & Ruback, R. B. (2012). Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*, 1003–1016. <https://doi.org/10.1109/TPAMI.2011.176>
- Geen, R. G. (1990). *Human aggression*. Open University Press.
- Georgiev, A. V., Klimczuk, A. C., Traficonte, D. M., & Maestriperi, D. (2013). When violence pays: A cost-benefit analysis of aggressive behavior in animals and humans. *Evolutionary Psychology*, *11*, 147470491301100320.

- Gerrard, G., & Thompson, R. (2011). Two million cameras in the UK. *CCTV Image*, 42, 2.
- Gill, M., & Spriggs, A. (2005). *Assessing the impact of CCTV*. Home Office Research, Development and Statistics Directorate London. Retrieved from <https://www.cctvusergroup.com/downloads/file/Martin%20gill.pdf>
- Gintis, H. (2010). Social norms as choreography. *Politics, Philosophy & Economics*, 9, 251–264. <https://doi.org/10.1177/1470594X09345474>
- Goldstein, J. (1975). *Aggression and crimes of violence*. Oxford University Press.
- Goudriaan, H., & Nieuwbeerta, P. (2007). Contextual determinants of juveniles' willingness to report crimes. *Journal of Experimental Criminology*, 3, 89–111. <https://doi.org/10.1007/s11292-007-9030-4>
- Graham, K., & Homel, R. (2012). *Raising the bar*. Routledge.
- Graham, K., Osgood, W., Zibrowski, E., Purcell, J., Gliksman, L., Leonard, K., ... Toomey, T. (2004). The effect of the Safer Bars programme on physical aggression in bars: Results of a randomized controlled trial. *Drug and Alcohol Review*, 23, 31–41. <https://doi.org/10.1080/09595230410001645538>
- Graham, K., Tremblay, P. F., Wells, S., Pernanen, K., Purcell, J., & Jelley, J. (2006). Harm, intent, and the nature of aggressive behavior: Measuring naturally occurring aggression in bar room settings. *Assessment*, 13, 280–296. <https://doi.org/10.1177/1073191106288180>
- Graham, K., & Wells, S. (2003). 'Somebody's gonna get their head kicked in tonight!': aggression among young males in bars—a question of values? *The British Journal of Criminology*, 43, 546–566.
- Gregoski, M., Malone, W. A., & Richardson, D. S. (2005). Measuring direct and indirect aggression: Is there a response bias? *Psychological Reports*, 97, 563–566.
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., ... DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models.

- Ecological Modelling*, 198(1–2), 115–126.
<https://doi.org/10.1016/j.ecolmodel.2006.04.023>
- Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, 221, 2760–2768.
<https://doi.org/10.1016/j.ecolmodel.2010.08.019>
- Groff, E. R. (2014). Informal social control and crime events. *Journal of Contemporary Criminal Justice*, 31, 90–106. <https://doi.org/10.1177/1043986214552619>
- Hadfield, P., Lister, S., & Traynor, P. (2009). ‘This town’s a different town today’: Policing and regulating the night-time economy. *Criminology and Criminal Justice*, 9, 465–485. <https://doi.org/10.1177/1748895809343409>
- Hair, J. F. (1998). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Haslam, A. S. (2004). *Psychology in organizations: The social identity approach*. SAGE Publications Inc.
- Haslam, A. S., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin*, 25, 809–818.
- Haslam, A. S., Reicher, S. D., & Levine, M. (2012). When other people are heaven, when other people are hell: How social identity determines the nature and impact of social support. In J. Jetten, C. Haslam, & A. S. Haslam (Eds.), *The social cure: Identity, health and well-being* (pp. 157–174). London; New York: Psychology Press.
- Haslam, A. S., & Turner, J. C. (1992). Context-dependent variation in social stereotyping 2: The relationship between frame of reference, self-categorization and accentuation. *European Journal of Social Psychology*, 22, 251–277.
<https://doi.org/10.1002/ejsp.2420220305>

- Haslam, A. S., & Turner, J. C. (1995). Context-dependent variation in social stereotyping 3: Extremism as a self-categorical basis for polarized judgement. *European Journal of Social Psychology*, 25, 341–371. <https://doi.org/10.1002/ejsp.2420250307>
- Haslam, A. S., Turner, J. C., Oakes, P. J., McGarty, C., & Reynolds, K. J. (1997). The group as a basis for emergent stereotype consensus. *European Review of Social Psychology*, 8, 203–239. <https://doi.org/10.1080/14792779643000128>
- Heinke, D., Carslaw, G., & Christian, J. (2013). An agent-based simulation of destigmatization (DSIM): Introducing a contact theory and self-fulfilling prophecy approach. *Journal of Artificial Societies and Social Simulation*, 16.
- Henry, A. D., Prałat, P., & Zhang, C.-Q. (2011). Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108, 8605–8610. <https://doi.org/10.1073/pnas.1014486108>
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604. <https://doi.org/10.1146/annurev.psych.53.100901.135109>
- Hobbs, D. (2003). *Bouncers: Violence and Governance in the Night-time Economy*. Oxford University Press.
- Hobbs, D., Hadfield, P., Lister, S., & Winlow, S. (2003). *Bouncers: violence and governance in the night time economy*. Oxford, UK: Oxford University Press.
- Hobbs, D., Hadfield, P., Lister, S., & Winlow, S. (2005). Violence and control in the night-time economy. *European Journal of Crime, Criminal Law and Criminal Justice*, 13, 89–102. <https://doi.org/10.1163/1571817053558310>
- Hobbs, D., Lister, S., Hadfield, P., Winlow, S., & Hall, S. (2000). Receiving shadows: Governance and liminality in the night-time economy. *The British Journal of Sociology*, 51, 701–717. <https://doi.org/10.1080/00071310020015334>

- Hochstetler, A., Copes, H., & Forsyth, C. J. (2014). The fight: Symbolic expression and validation of masculinity in working class tavern culture. *American Journal of Criminal Justice*, 39, 493–510. <https://doi.org/10.1007/s12103-013-9222-6>
- Hogg, M. A. (2004). Uncertainty and extremism: Identification with high entitativity groups under conditions of uncertainty. In V. Yzerbyt, C. M. Judd, & O. Corneille (Eds.), *The psychology of group perception: Perceived variability, entitativity, and essentialism* (pp. 401–418). Philadelphia, PA: Psychology Press.
- Hogg, M. A. (2006). Social Identity Theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (1st ed.). Stanford, Calif: Stanford Social Sciences.
- Hogg, M. A. (2012). Uncertainty-identity theory. In P. A. M. V. Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology: Volume two*. SAGE Publications.
- Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, 16, 7–30. <https://doi.org/10.1111/j.1468-2885.2006.00003.x>
- Hogg, M. A., & Terry, D. I. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review*, 25, 121–140. <https://doi.org/10.5465/AMR.2000.2791606>
- Hollis, M. E., Felson, M., & Welsh, B. C. (2013). The capable guardian in routine activities theory: A theoretical and conceptual reappraisal. *Crime Prevention and Community Safety*, 15, 65–79. <https://doi.org/10.1057/cpcs.2012.14>
- Homel, R., Clark, J., & others. (1994). The prediction and prevention of violence in pubs and clubs. *Crime Prevention Studies*, 3, 1–46.
- Homel, R., & Tomsen, S. (1993). Hot spots for violence: The environment of pubs and clubs. *Australian Institute of Criminology*, 22, 679–697.

- Huesmann, R. L. (1988). An information processing model for the development of aggression. *Aggressive Behavior, 14*, 13–24. [https://doi.org/10.1002/1098-2337\(1988\)14:1<13::AID-AB2480140104>3.0.CO;2-J](https://doi.org/10.1002/1098-2337(1988)14:1<13::AID-AB2480140104>3.0.CO;2-J)
- Huesmann, R. L. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E. D. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy*. Elsevier.
- Huesmann, R. L., & Taylor, L. D. (2006). The role of media violence in violent behavior. *Annual Review of Public Health, 27*, 393–415. <https://doi.org/10.1146/annurev.publhealth.26.021304.144640>
- Hughes, K., Bellis, M. A., Leckenby, N., Quigg, Z., Hardcastle, K., Sharples, O., & Llewellyn, D. J. (2014). Does legislation to prevent alcohol sales to drunk individuals work? Measuring the propensity for night-time sales to drunks in a UK city. *Journal of Epidemiology and Community Health, 68*, 453–456. <https://doi.org/10.1136/jech-2013-203287>
- Iannaccone, L. R., & Makowsky, M. D. (2007). Accidental atheists? Agent-based explanations for the persistence of religious regionalism. *Journal for the Scientific Study of Religion, 46*, 1–16. <https://doi.org/10.1111/j.1468-5906.2007.00337.x>
- Insko, C. A., Schopler, J., Gaertner, L., Wildschut, T., Kozar, R., Pinter, B., ... Montoya, M. R. (2001). Interindividual–intergroup discontinuity reduction through the anticipation of future interaction. *Journal of Personality and Social Psychology, 80*, 95–111. <https://doi.org/10.1037/0022-3514.80.1.95>
- Insko, C. A., Schopler, J., Hoyle, R. H., Dardis, G. J., & Graetz, K. A. (1990). Individual–group discontinuity as a function of fear and greed. *Journal of Personality and Social Psychology, 58*, 68–79. <https://doi.org/10.1037/0022-3514.58.1.68>

- Jackson-Jacobs, C. (2004). Taking a beating: The narrative gratifications of fighting as an underdog. In J. Ferrell, K. Hayward, W. Morrison, & M. Presdee (Eds.), *Cultural criminology unleashed*. Routledge.
- Jankowski, M. S. (1991). *Islands in the street: Gangs and American urban society*. University of California Press.
- Janowitz, M. (1975). Sociological theory and social control. *American Journal of Sociology*, 82–108.
- Jans, L., Leach, C. W., Garcia, R. L., & Postmes, T. (2015). The development of group influence on in-group identification: A multilevel approach. *Group Processes & Intergroup Relations*, 18(2), 190–209. <https://doi.org/10.1177/1368430214540757>
- Johnson, A., & Sackett, R. (1998). Direct systematic observation of behavior. In H. R. Bernard (Ed.), *Handbook of methods in cultural anthropology* (pp. 301–332). AltaMira Press.
- Kaufmann, H. (1965). Definitions and methodology in the study of aggression. *Psychological Bulletin*, 64, 351–364. <https://doi.org/10.1037/h0022530>
- Keiser, R. L. (1969). *The vice lords, warriors of the streets*. Holt, Rinehart and Winston.
- Keltner, D., Young, A. C., & Buswell, B. N. (1997). Appeasement in human emotion, social practice, and personality. *Aggressive Behavior*, 359–374.
- Kemeny, M. E., Gruenewald, T. L., & Dickerson, S. S. (2004). Shame as the emotional response to threat to the social self: Implications for behavior, physiology, and health. *Psychological Inquiry*, 15, 153–160.
- Klein, Katherine J., Bliese, P. D., J, W., Dansereau, F., Gavin, M. B., Griffin, M. A., ... Bligh, M. C. (2000). Multilevel analytical techniques: Commonalities, differences, and continuing questions. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 512–553). San Francisco, CA, US: Jossey-Bass.

- Klein, O., & Azzi, A. E. (2001). The strategic confirmation of meta-stereotypes: How group members attempt to tailor an out-group's representation of themselves. *British Journal of Social Psychology, 40*, 279–293.
- Knuijver, S., Kornfeld, A., Ravensbergen, L., Timmer, C., & Witteman, J. (2006). *Aggression and social identities* (Unpublished Bachelor Thesis). University of Groningen, The Netherlands.
- Koudenburg, N., Postmes, T., Gordijn, E. H., & van Mourik Broekman, A. (2015). Uniform and complementary social interaction: Distinct pathways to solidarity. *PLOS ONE, 10*. <https://doi.org/10.1371/journal.pone.0129061>
- Krakauer, D. C., Page, K., & Flack, J. C. (2011). The immuno-dynamics of conflict intervention in social systems. *PloS One, 6*. <https://doi.org/10.1371/journal.pone.0022709>
- Kramer, R. M., & Cook, K. S. (Eds.). (2004). *Trust and distrust in organizations: Dilemmas and approaches*. Russell Sage Foundation.
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. Boston: McGraw-Hill/Irwin.
- Laney, C., & Takarangi, M. K. T. (2013). False memories for aggressive acts. *Acta Psychologica, 143*, 227–234. <https://doi.org/10.1016/j.actpsy.2013.04.001>
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York, NY: Appleton-Century-Croft.
- Le Bon, G. (1960). *The crowd: A study of the popular mind*. Viking Press. (Original work published 1897)
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815–852. <https://doi.org/10.1177/1094428106296642>

- LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, *52*, 31–60.
<https://doi.org/10.3102/00346543052001031>
- Leombruni, R., & Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and Its Applications*, *355*, 103–109.
<https://doi.org/10.1016/j.physa.2005.02.072>
- Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory: a framework for social identity, social cognition, and intergroup relations. In M. P. Z. and J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 43, pp. 63–113). Academic Press.
- Levine, M. (1999). Rethinking bystander nonintervention: Social categorization and the evidence of witnesses at the James Bulger murder trial. *Human Relations*, *52*, 1133–1155. <https://doi.org/10.1177/001872679905200902>
- Levine, M., Cassidy, C., & Jentsch, I. (2010). The implicit identity effect: Identity primes, group size, and helping. *British Journal of Social Psychology*, *49*(4), 785–802.
<https://doi.org/10.1348/014466609X480426>
- Levine, M., & Crowther, S. (2008). The responsive bystander: How social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology*, *95*, 1429–1439. <https://doi.org/10.1037/a0012634>
- Levine, M., Lowe, R., Best, R., & Heim, D. (2012). ‘We police it ourselves’: Group processes in the escalation and regulation of violence in the night-time economy: Night-time economy violence. *European Journal of Social Psychology*, *42*, 924–932.
<https://doi.org/10.1002/ejsp.1905>
- Levine, M., & Manning, R. (2013). Social identity, group processes, and helping in emergencies. *European Review of Social Psychology*, *24*, 225–251.
<https://doi.org/10.1080/10463283.2014.892318>

- Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*, 443–453.
<https://doi.org/10.1177/0146167204271651>
- Levine, M., Taylor, P. J., & Best, R. (2011). Third parties, violence, and conflict resolution: The role of group size and collective action in the microregulation of violence. *Psychological Science*, *22*, 406–412. <https://doi.org/10.1177/0956797611398495>
- Leyens, J.-P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review*, *4*, 186–197. https://doi.org/10.1207/S15327957PSPR0402_06
- Liebst, L. S. (2017). *Consolation in the aftermath of assaults in public places: A multilevel case-control study based on naturally occurring events*. Manuscript in preparation.
- Liebst, L. S., Heinskou, M. B., & Ejbye-Ernst, P. (2017). On the actual risk of bystander intervention. *Journal of Research in Crime and Delinquency*.
- Lindgaard, M. R., & Copes, H. (2017). Observational methods of offender decision making. In W. Bernasco, H. Elffers, & J.-L. van Gelder (Eds.), *The Oxford handbook of offender decision making*. Oxford University Press.
- Lindgaard, M. R., Liebst, L. S., Bernasco, W., Heinskou, M. B., Philpot, R., Levine, M., & Verbeek, P. (2017). Consolation in the aftermath of robberies resembles post-aggression consolation in chimpanzees. *PLOS ONE*, *12*.
<https://doi.org/10.1371/journal.pone.0177725>
- Lister, S. (2009). Police and policing. In A. Hucklesby & A. Wahidin (Eds.), *Criminal justice* (1 edition). Oxford; New York: OUP Oxford.
- Little, T. D., Rodkin, P. C., & Hawley, P. H. (2013). *Aggression and adaptation: The bright side to bad behavior*. Routledge.

- Littman, R., & Paluck, E. L. (2015). The cycle of violence: Understanding individual participation in collective violence. *Political Psychology, 36*, 79–99.
<https://doi.org/10.1111/pops.12239>
- Livingstone, A., & Haslam, A. S. (2008). The importance of social identity content in a setting of chronic social conflict: Understanding intergroup relations in Northern Ireland. *British Journal of Social Psychology, 47*, 1–21.
<https://doi.org/10.1348/014466607X200419>
- Lorenz, K. (1966). *On aggression*. London: Methuen & Co, Ltd.
- Loveday, B. (2005). The 2003 licensing act: Alcohol use and anti-social behaviour in England and Wales. *The Police Journal, 78*, 191–208.
<https://doi.org/10.1350/pojo.2005.78.3.191>
- Lowe, R. D., Levine, M., Best, R. M., & Heim, D. (2012). Bystander reaction to women fighting: Developing a theory of intervention. *Journal of Interpersonal Violence, 27*, 1802–1826.
- Macintyre, S., & Homel, R. (1997). Danger on the dance floor: A study of interior design, crowding and aggression in nightclubs. *Policing for Prevention: Reducing Crime, Public Intoxication and Injury, 7*, 91–113.
- Mackie, D., & Cooper, J. (1984). Attitude polarization: Effects of group membership. *Journal of Personality and Social Psychology, 46*, 575–585. <https://doi.org/10.1037/0022-3514.46.3.575>
- Mackie, D. M. (1986). Social identification effects in group polarization. *Journal of Personality and Social Psychology, 50*, 720–728. <https://doi.org/10.1037/0022-3514.50.4.720>
- MacLean, P. D. (1990). *The triune brain in evolution: Role in paleocerebral functions*. Springer Science & Business Media.

- MacLennan, B. J. (2007). NetLogo flocking model. Retrieved 4 June 2017, from <http://web.eecs.utk.edu/~mclennan/Classes/420-594-F07/NetLogo/Flock.html>
- Manzo, G. (2014). Potentialités et limites de la simulation multi-agents: Une introduction. *Revue Française de Sociologie*, 55, 653–688.
- Marques, J. M., Abrams, D., Páez, D., & Hogg, M. A. (2001). Social categorization, social identification, and rejection of deviant group members. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 400–424). Blackwell Publishers Ltd. <https://doi.org/10.1002/9780470998458.ch17>
- Marsh, P., & Fox-Kibby, K. (1992). *Drinking and public disorder* (Report). London: MCM Research. Retrieved from <http://www.sirc.org/publik/dandpd.pdf>
- Marsh, P., Rosser, E., & Harré, R. (1978). *The rules of disorder*. Routledge & Kegan Paul.
- Martens, J. P., Tracy, J. L., & Shariff, A. F. (2012). Status signals: Adaptive benefits of displaying and observing the nonverbal expressions of pride and shame. *Cognition and Emotion*, 26, 390–406. <https://doi.org/10.1080/02699931.2011.645281>
- McGarty, C., Turner, J. C., Hogg, M. A., David, B., & Wetherell, M. S. (1992). Group polarization as conformity to the prototypical group member. *British Journal of Social Psychology*, 31, 1–19. <https://doi.org/10.1111/j.2044-8309.1992.tb00952.x>
- McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review*, 4, 95–105. https://doi.org/10.1207/S15327957PSPR0401_8
- Moore, C. W. (2014). *The mediation process: practical strategies for resolving conflict*. John Wiley & Sons.
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37, 17–23. <https://doi.org/10.2307/2332142>

- 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*, 103–122. <https://doi.org/10.1002/ejsp.2420220202>
- Muller, M. N. (2002). Agonistic relations among Kanyawara chimpanzees. In C. Boesch, G. Hohmann, & L. Marchant (Eds.), *Behavioural Diversity in Chimpanzees and Bonobos* (pp. 112–124). Cambridge University Press.
- Murakami, H. (1994). *Dance dance dance*. New York: Kodansha America Inc.
- Murphy, N. (2016). What we know and the future of accuracy research. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 404–424). Cambridge, United Kingdom: Cambridge University Press.
- Myers, R. H. (1990). *Classical and Modern Regression with Applications* (2nd ed.). Boston, MA: Duxbury/Thompson Learning.
- Nakamura, R., Suhrcke, M., Pechey, R., Morciano, M., Roland, M., & Marteau, T. M. (2014). Impact on alcohol purchasing of a ban on multi-buy promotions: a quasi-experimental evaluation comparing Scotland with England and Wales: Scottish alcohol multi-buy ban evaluation. *Addiction*, *109*, 558–567. <https://doi.org/10.1111/add.12419>
- Nassauer, A., & Legewie, N. M. (2016). Visual data analysis: Towards a methodological framework for a novel trend in studying behavior. Presented at the 111th Annual meeting of the American Sociological Association, Washington, USA.
- Newton, A. (2015). Crime and the NTE: Multi-classification crime (MCC) hot spots in time and space. *Crime Science*, *4*. <https://doi.org/10.1186/s40163-015-0040-7>
- Newton, A., & Hirschfield, A. (2009). Measuring violence in and around licensed premises: The need for a better evidence base. *Crime Prevention and Community Safety*, *11*, 171–188. <https://doi.org/10.1057/cpcs.2009.12>

- Noldus, L. P. J. J. (1991). The Observer: A software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, & Computers*, *23*, 415–429. <https://doi.org/10.3758/BF03203406>
- Novelli, D., Drury, J., & Reicher, S. (2010). Come together: Two studies concerning the impact of group relations on personal space. *British Journal of Social Psychology*, *49*, 223–236. <https://doi.org/10.1348/014466609X449377>
- Nowak, A. (2017, March 23). *Computational models of interpersonal and social synchronization*. Presentation given at the International Convention of Psychological Science, Vienna, Austria.
- Oakes, P. J. (1996). The categorization process: Cognition and the group in the social psychology of stereotyping. In W. P. Robinson & H. Tajfel (Eds.), *Social groups and identities: Developing the legacy of Henri Tajfel*. Psychology Press.
- Oakes, P. J., Haslam, A. S., & Turner, J. C. (1994). *Stereotyping and social reality* (Vol. xiv). Malden: Blackwell Publishing.
- Otten, S. (2009). Social categorization, intergroup emotions and aggressive interactions. In S. Otten, K. Sassenberg, & T. Kessler (Eds.), *Intergroup relations: The role of motivation and emotion*. Hove ; New York: Psychology Press.
- Otten, S., & Jacoby, J. (2004). *Perceived aggression by Dutch and German soccer supporters during the European championships 2004*. Unpublished data, The Netherlands.
- Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, *36*, 77–89. <https://doi.org/10.1006/jesp.1999.1399>
- Parks, M. J., Osgood, D. W., Felson, R. B., Wells, S., & Graham, K. (2013). Third party involvement in barroom conflicts: Third parties in barroom conflicts. *Aggressive Behavior*, *39*, 257–268. <https://doi.org/10.1002/ab.21475>

- Paternotte, C., & Grose, J. (2013). Social norms and game theory: Harmony or discord? *The British Journal for the Philosophy of Science*, *64*, 551–587.
<https://doi.org/10.1093/bjps/axs024>
- Pedersen, W., Copes, H., & Sandberg, S. (2016). Alcohol and violence in nightlife and party settings: A qualitative study: Alcohol, violence and nightlife. *Drug and Alcohol Review*, *35*, 557–563. <https://doi.org/10.1111/dar.12395>
- Pinto, I. R., Marques, J. M., Levin, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of Personality and Social Psychology*, *99*, 107–119. <https://doi.org/10.1037/a0018187>
- Planty, M. (2002). *Third-party involvement in violent crime, 1993-99* (BJS Special Report). US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Retrieved from www.ojp.usdoj.gov/bjs
- Porter, T. (2015). *Research on violence reduction* (Report). London: Cragg Ross Dawson. Retrieved from <https://www.sia.homeoffice.gov.uk/Documents/research/sia-violence-reduction-report.pdf>
- Portney, L. G., & Watkins, M. P. (2009). Reliability of measurements. In *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, N.J: Prentice Hall.
- Postmes, T., Haslam, A. S., & Swaab, R. I. (2005). Social influence in small groups: An interactive model of social identity formation. *European Review of Social Psychology*, *16*, 1–42. <https://doi.org/10.1080/10463280440000062>
- Postmes, T., & Spears, R. G. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, *123*, 238–259. <https://doi.org/10.1037/0033-2909.123.3.238>
- Prentice, D. A., Miller, D. T., & Lightdale, J. R. (1994). Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond

- groups. *Personality and Social Psychology Bulletin*, 20, 484–493.
<https://doi.org/10.1177/0146167294205005>
- Railsback, S. F., & Grimm, V. (2011). *Agent-based and individual-based modeling: A practical introduction*. Princeton University Press.
- Reicher, S. D. (1984). The St. Pauls' riot: An explanation of the limits of crowd action in terms of a social identity model. *European Journal of Social Psychology*, 14, 1–21.
<https://doi.org/10.1002/ejsp.2420140102>
- Reicher, S. D. (1987). Crowd behaviour as social action. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory* (Vol. x). Cambridge, MA, US: Basil Blackwell.
- Reicher, S. D. (1996). 'The Battle of Westminster': developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict. *European Journal of Social Psychology*, 26, 115–134.
[https://doi.org/10.1002/\(SICI\)1099-0992\(199601\)26:1<115::AID-EJSP740>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1099-0992(199601)26:1<115::AID-EJSP740>3.0.CO;2-Z)
- Reicher, S. D. (2001). The psychology of crowd dynamics. In M. A. Hogg & S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 175–197). Oxford: Blackwell Publishers.
- Reicher, S. D., Cassidy, C., Wolpert, I., Hopkins, N., & Levine, M. (2006). Saving Bulgaria's Jews: an analysis of social identity and the mobilisation of social solidarity. *European Journal of Social Psychology*, 36(1), 49–72. <https://doi.org/10.1002/ejsp.291>
- Reicher, S. D., & Haslam, A. S. (2009). Beyond Help. In S. Stürmer & M. Snyder (Eds.), *The psychology of prosocial behavior* (pp. 289–309). Wiley-Blackwell.
<https://doi.org/10.1002/9781444307948.ch15>

- Reicher, S. D., & Hopkins, N. (2001). Psychology and the end of history: A critique and a proposal for the psychology of social categorization. *Political Psychology, 22*, 383–407. <https://doi.org/10.1111/0162-895X.00246>
- Reicher, S. D., Spears, R. G., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology, 6*, 161–198. <https://doi.org/10.1080/14792779443000049>
- Reicher, S. D., Stott, C., Cronin, P., & Adang, O. (2004). An integrated approach to crowd psychology and public order policing. *Policing: An International Journal of Police Strategies & Management, 27*, 558–572. <https://doi.org/10.1108/13639510410566271>
- Reiss, A. J. (1951). Delinquency as the failure of personal and social controls. *American Sociological Review, 16*, 196. <https://doi.org/10.2307/2087693>
- Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies, 28*, 377–408. <https://doi.org/10.1177/0170840607076007>
- Reynald, D. M. (2017). Informal guardians and offender decision making. In W. Bernasco, H. Elffers, & J.-L. van Gelder (Eds.), *The Oxford handbook of offender decision making*. Oxford University Press.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological), 39*, 172–212.
- Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*, 407–419. <https://doi.org/10.1002/ab.20066>
- Roberts, M. (2006). From ‘creative city’ to ‘no-go areas’ – The expansion of the night-time economy in British town and city centres. *Cities, 23*, 331–338. <https://doi.org/10.1016/j.cities.2006.05.001>
- Rogerson, P. (2001). *Statistical Methods for Geography*. London: SAGE Publications.

- Romero, T. (forthcoming). The peacemaking and peacekeeping role of chimpanzee bystanders. In B. Peters & P. Verbeek (Eds.), *Peace ethology: Behavioral processes and systems of peace*. Chichester, UK: John Wiley & Sons Inc.
- Rummel, R. J. (1979). *Understanding conflict and war, Vol. 2: The conflict helix* (2nd ed.). Beverly Hills: SAGE Publications.
- Sakamaki, T., & Nakamura, M. (2015). Intergroup relationships. In M. Nakamura, K. Hosaka, N. Itoh, & K. Zamma (Eds.), *Mahale Chimpanzees: 50 Years of Research* (pp. 128–140). Cambridge: Cambridge University Press.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global sensitivity analysis: The primer*. John Wiley & Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons.
- Salzarulo, L. (2004). *Formalizing self-categorization theory to simulate the formation of social groups*. Unpublished paper presented at the Proceedings of the 2nd International Conferences of the European Social Simulation Association, Valladolid, Spain.
- Salzarulo, L. (2006). A continuous opinion dynamics model based on the principle of meta-contrast. *Journal of Artificial Societies and Social Simulation*, 9. Retrieved from <http://jasss.soc.surrey.ac.uk/9/1/13.html>
- Sanders, W. B. (1994). *Gangbangs and drive-bys: Grounded culture and juvenile gang violence*. Aldine de Gruyter.
- Saunders, D. G. (1991). Procedures for adjusting self-reports of violence for social desirability bias. *Journal of Interpersonal Violence*, 6, 336–344.
<https://doi.org/10.1177/088626091006003006>
- Schellenberg, R. C., Parks-Savage, A., & Rehfuss, M. (2007). Reducing levels of elementary school violence with peer mediation. *Professional School Counseling*, 10, 475–481.
<https://doi.org/10.5330/prsc.10.5.q786607713v5q044>

- Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, *1*, 143–186. <https://doi.org/10.1080/0022250X.1971.9989794>
- Schrieff, L., Tredoux, C., Dixon, J., & Finchilescu, G. (2005). Patterns of racial segregation in university residence dining-halls. *South African Journal of Psychology*, *35*, 433–443. <https://doi.org/10.1177/008124630503500303>
- Scott, J. P. (1958). *Aggression*. University of Chicago Press.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Sicotte, P. (1995). Interpositions in conflicts between males in bimale groups of mountain gorillas. *Folia Primatologica*, *65*, 14–24. <https://doi.org/10.1159/000156871>
- Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLOS ONE*, *8*. <https://doi.org/10.1371/journal.pone.0052766>
- Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*, 17228–17233. <https://doi.org/10.1073/pnas.0706741104>
- Slotter, E. B., & Finkel, E. J. (2011). I³ theory: Instigating, impelling, and inhibiting factors in aggression. In P. R. Shaver & M. Mikulincer (Eds.), *Human aggression and violence: Causes, manifestations, and consequences* (pp. 35–52). Washington, DC, US: American Psychological Association.
- Slutkin, G., Ransford, C., Decker, B., & Volker, K. (2015). *Cure violence—An evidence based method to reduce shootings and killings*. Washington, DC. Retrieved from <http://www.gppplatform.ch/sites/default/files/Cure%20Violence.pdf>
- Smaldino, P. E., Calanchini, J., & Pickett, C. L. (2015). Theory development with agent-based models. *Organizational Psychology Review*, *5*, 300–317. <https://doi.org/10.1177/2041386614546944>

- Smaldino, P. E., Pickett, C., Sherman, J., & Schank, J. (2012). An agent-based model of social identity dynamics. *Journal of Artificial Societies and Social Simulation*, *15*, 1–7.
- Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, *11*, 87–104. <https://doi.org/10.1177/1088868306294789>
- Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin*, *22*, 635–642. <https://doi.org/10.1177/0146167296226008>
- Smith, J. R., & Hogg, M. A. (2008). Social identity and attitudes. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (1st ed.). New York; London: Psychology Press.
- Smith, L. G. E., & Postmes, T. (2011a). Shaping stereotypical behaviour through the discussion of social stereotypes: Shaping stereotypical behaviour. *British Journal of Social Psychology*, *50*, 74–98. <https://doi.org/10.1348/014466610X500340>
- Smith, L. G. E., & Postmes, T. (2011b). The power of talk: Developing discriminatory group norms through discussion: The power of talk. *British Journal of Social Psychology*, *50*, 193–215. <https://doi.org/10.1348/014466610X504805>
- Stott, C., Adang, O., Livingstone, A., & Schreiber, M. (2007). Variability in the collective behaviour of England fans at Euro2004: ‘Hooliganism’, public order policing and social change. *European Journal of Social Psychology*, *37*, 75–100. <https://doi.org/10.1002/ejsp.338>
- Stott, C., Adang, O., Livingstone, A., & Schreiber, M. (2008). Tackling football hooliganism: A quantitative study of public order, policing and crowd psychology. *Psychology, Public Policy, and Law*, *14*, 115–141. <https://doi.org/10.1037/a0013419>
- Stott, C., Hutchison, P., & Drury, J. (2001). ‘Hooligans’ abroad? Inter-group dynamics, social identity and participation in collective ‘disorder’ at the 1998 World Cup Finals. *British Journal of Social Psychology*, *40*, 359–384. <https://doi.org/10.1348/014466601164876>

- Stott, C., Scothern, M., & Gorringer, H. (2013). Advances in liaison based public order policing in England: Human rights and negotiating the management of protest? *Policing*, 7, 212–226. <https://doi.org/10.1093/police/pat007>
- Stott, C., West, O., & Radburn, M. (2016). Policing football ‘risk’? A participant action research case study of a liaison-based approach to ‘public order’. *Policing and Society*, 1–16. <https://doi.org/10.1080/10439463.2015.1126267>
- Struch, N., & Schwartz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from in-group bias. *Journal of Personality and Social Psychology*, 56, 364–373. <https://doi.org/10.1037/0022-3514.56.3.364>
- Stürmer, S., & Snyder, M. (2009). Helping ‘Us’ versus ‘Them’. In S. Stürmer & M. Snyder (Eds.), *The psychology of prosocial behavior* (pp. 33–58). 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, 943–956. <https://doi.org/10.1177/0146167206287363>
- Suvilehto, J. T., Glerean, E., Dunbar, R. I. M., Hari, R., & Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences*, 112, 13811–13816. <https://doi.org/10.1073/pnas.1519231112>
- Swann, W. B., & Jetten, J. (2017). Restoring agency to the human actor. *Perspectives on Psychological Science*, 12, 382–399. <https://doi.org/10.1177/1745691616679464>
- Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119, 441–456. <https://doi.org/10.1037/a0028589>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston, MA: Pearson Education.

- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149–178.
<https://doi.org/10.1002/ejsp.2420010202>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations, 33*, 74.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In W. G. Austin & S. Worchel (Eds.), *Psychology of intergroup relations*. Nelson-Hall Publishers.
- Tajima, T., & Kurotori, H. (2010). Nonaggressive interventions by third parties in conflicts among captive Bornean orangutans (*Pongo pygmaeus*). *Primates, 51*, 179–182.
<https://doi.org/10.1007/s10329-009-0180-z>
- Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology, 48*, 427–446.
<https://doi.org/10.1348/014466608X373589>
- Taylor, P. J., Donald, I., & Conchie, S. M. (2012). *The triple-interact as a building block of negotiation* (SSRN Scholarly Paper). Rochester, NY: Social Science Research Network.
- Tedeschi, J. T., & Felson, R. B. (1994). *Violence, aggression, and coercive actions* (Vol. xii). Washington, DC, US: American Psychological Association.
<https://doi.org/10.1037/10160-000>
- Thrasher, J., & Vallier, K. (2015). The fragility of consensus: Public reason, diversity and stability: The fragility of consensus. *European Journal of Philosophy, 23*, 933–954.
<https://doi.org/10.1111/ejop.12020>
- Tipping, P., & Lloyd, D. (2015). *Continuity with change – the implications of CSR15* (Report). Prepared for the Police and Crime Commissioners Association.

- Tolan, P. H. (2007). Understanding violence. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (1st ed.). New York: Cambridge University Press.
- Tolan, P. H., Gorman-Smith, D., & Henry, D. (2006). Family violence. *Annual Review of Psychology, 57*, 557–583. <https://doi.org/10.1146/annurev.psych.57.102904.190110>
- Tomsen, S. (1997). A top night: Social protest, masculinity and the culture of drinking violence. *The British Journal of Criminology, 37*, 90–102.
<https://doi.org/10.1093/oxfordjournals.bjc.a014152>
- Townsley, M., & Grimshaw, R. (2013). The consequences of queueing: Crowding, situational features and aggression in entertainment precincts. *Crime Prevention & Community Safety, 15*, 23–47. <https://doi.org/10.1057/cpcs.2012.12>
- Turner, J. C. (1991). *Social Influence*. Brooks/cole Publishing Company.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory* (Vol. x). Cambridge, MA, US: Basil Blackwell.
- Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology, 25*(3), 237–252. <https://doi.org/10.1111/j.2044-8309.1986.tb00732.x>
- van Bommel, M., van Prooijen, J.-W., Elffers, H., & van Lange, P. A. M. (2014). Intervene to be seen: The power of a camera in attenuating the bystander effect. *Social Psychological and Personality Science, 5*, 459–466.
<https://doi.org/10.1177/1948550613507958>
- van Gelder, J.-L. van, Otte, M., & Luciano, E. C. (2014). Using virtual reality in criminological research. *Crime Science, 3*, 10. <https://doi.org/10.1186/s40163-014-0010-5>

- van Rooy, D. (2012). A connectionist ABM of social categorization processes. *Advances in Complex Systems, 15*, 1250077. <https://doi.org/10.1142/S0219525912500774>
- Vasquez, E. A., Wenborne, L., Peers, M., Alleyne, E., & Ellis, K. (2015). Any of them will do: In-group identification, out-group entitativity, and gang membership as predictors of group-based retribution: Gangs and group-based retaliation. *Aggressive Behavior, 41*, 242–252. <https://doi.org/10.1002/ab.21581>
- Vidmar, N. (2001). Retribution and revenge. In J. Sanders & V. L. Hamilton (Eds.), *Handbook of justice research in law* (pp. 31–63). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/0-306-47379-8_2
- Viki, T. G., & Abrams, D. (2013). The social influence of groups on individuals. In J. L. Wood & T. A. Gannon (Eds.), *Crime and crime reduction: The importance of group processes*. Routledge.
- von Rohr, C. R., Koski, S. E., Burkart, J. M., Caws, C., Fraser, O. N., Ziltener, A., & van Schaik, C. P. (2012). Impartial third-party interventions in captive chimpanzees: A reflection of community concern. *PLoS One, 7*. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0032494>
- von Sivers, I., Templeton, A., Köster, G., Drury, J., & Philippides, A. (2014). Humans do not always act selfishly: Social identity and helping in emergency evacuation simulation. *Transportation Research Procedia, 2*, 585–593. <https://doi.org/10.1016/j.trpro.2014.09.099>
- Vorauer, J. D., Main, K. J., Roy, S. A., & Hunter, A. J. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology, 78*, 690–707. <https://doi.org/10.1037/0022-3514.78.4.690>
- Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting reliable information in investigative interviews. *Policy Insights from the Behavioral and Brain Sciences, 1*, 129–136.

- Warton, D. I., & Hui, F. K. C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, *92*, 3–10. <https://doi.org/10.1890/10-0340.1>
- Weenink, D. (2014). Frenzied attacks. A micro-sociological analysis of the emotional dynamics of extreme youth violence: Frenzied attacks. *The British Journal of Sociology*, *65*, 411–433. <https://doi.org/10.1111/1468-4446.12088>
- Weisburd, D., Groff, E., & Yang, S.-M. (2012). *The criminology of place: Street segments and our understanding of the crime problem*. Oxford; New York: Oxford University Press.
- Wells, S., & Graham, K. (1999). The frequency of third-party involvement in incidents of barroom aggression. *Contemporary Drug Problems*, *26*, 457–480.
- Welsh, B. C., & Farrington, D. P. (2002). *Crime prevention effects of closed circuit television: a systematic review*. London: Home Office Research, Development and Statistics Directorate.
- Welsh, B. C., & Farrington, D. P. (2009). Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Quarterly*, *26*, 716–745. <https://doi.org/10.1080/07418820802506206>
- Wilensky, U. (1999). NetLogo (Version 5.0.2) [Windows 7]. Evanston, IL: Northwestern University, Center for Connected Learning and Computer-Based Modeling. Retrieved from <http://ccl.northwestern.edu/netlogo>
- Wilson, M. A., Bradford, E., & Lemanski, L. (2009). The Role of group processes in terrorism. In S. Otten, K. Sassenberg, & T. Kessler (Eds.), *Intergroup relations: The role of motivation and emotion (A Festschrift for Amélie Mummendey)*. Psychology Press.
- Wilson, M. L., & Wrangham, R. W. (2003). Intergroup relations in chimpanzees. *Annual Review of Anthropology*, *32*, 363–392. <https://doi.org/10.1146/annurev.anthro.32.061002.120046>
- Winlow, S., & Hall, S. (2006). *Violent night: Urban leisure and contemporary culture*. Berg.

- Wittig, R. M., Crockford, C., Langergraber, K. E., & Zuberbühler, K. (2014). Triadic social interactions operate across time: A field experiment with wild chimpanzees. *Proc. R. Soc. B*, *281*, 20133155. <https://doi.org/10.1098/rspb.2013.3155>
- Yuille, J. C., & Tollestrup, P. A. (1990). Some effects of alcohol on eyewitness memory. *Journal of Applied Psychology*, *75*, 268–273. <https://doi.org/10.1037/0021-9010.75.3.268>
- Zillmann, D. (1979). *Hostility and aggression* (1st ed.). Hillsdale: Psychology Press.
- Zillmann, D. (1988). Cognition-excitation interdependences in aggressive behavior. *Aggressive Behavior*, *14*, 51–64. [https://doi.org/10.1002/1098-2337\(1988\)14:1<51::AID-AB2480140107>3.0.CO;2-C](https://doi.org/10.1002/1098-2337(1988)14:1<51::AID-AB2480140107>3.0.CO;2-C)
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation*, *17*, 237–307.
- Zimbardo, P. G. (2008). *The Lucifer effect: Understanding how good people turn evil*. New York: Random House.

Appendix

Video Coding Book

Independent Variables

Variable	Categories for coding
Number of bystanders	<p>Number of people <u>NOT</u> including the protagonist and target who are actively involved in the incident</p> <p><i>Don't know option</i></p> <p>Please note, this will be coded post hoc using excel formulae</p>
Gender of people in fight / altercation	<p>Is each participant male or female?</p> <p>Code gender of all actively involved people</p> <p>Male</p> <p>Female</p> <p><i>Don't know option</i></p>
Quality / Severity of violence in the clip	<p>How serious is the violence?</p> <p>Mild (some posturing, aggressive gesturing, pushing and shoving, feinted punches, kicks or slaps but with no real conviction/or not landed on opponent)</p> <p>Moderate (fight takes place, numerous punches, slaps, kicks etc. landed but nobody badly hurt, knocked unconscious or kicked repeatedly)</p> <p>Severe (person badly hurt – bloody, knocked unconscious, laying on ground and hardly moving, kicked repeatedly/severely whilst on the floor)</p> <p>Other</p> <p><i>Don't know option</i></p>
Group Boundaries	<p>For post hoc coders:</p> <p>Is the conflict intragroup, intergroup or unknown?</p> <p>Intergroup (Conflict appears to be predominately between two or more different social groups)</p> <p>Intragroup (Conflict appears to be predominately between members of the same social group)</p> <p><i>Don't know option</i></p> <p>Please assign <i>all</i> individuals within the video a group ID number.</p> <p>If two or more individuals belong to the same group, please assign them with the same shared group number.</p> <p>If individuals belong to different groups, please ensure that they are assigned a separate (mutually exclusive) group number.</p> <p>To assess whether individuals belong to the same social group: Please note if the individuals coded appear to be cohesive and familiar with one another - e.g., by observing behaviours such as travelling around the streets together, walking in close proximity, stopping/starting at same time when walking together, accelerating/decelerating to keep up with one other, linking arms, wearing similar fashion or fancy dress, appearing to share friends and acquaintances determined via a shared focus of attention and close interpersonal distance. Please check full video several times, including time before and after conflict as individuals arrive/leave scene together</p>

Protagonist Events

The protagonist is the person driving the initial aggression towards another individual (known as the ‘target’).

If the protagonist performs any of the following acts towards another person, please press ‘q’ to start coding this protagonist’s action.

Event name	Code	Modifier	Rules and definitions
			<p>General Rules and definitions</p> <ol style="list-style-type: none"> 1. Timing: Events to be coded within 2 second time frame of real time (we may be a few tenths of a seconds out when pressing keys, which is OK) 2. Mutual exclusivity: Events are mutually exclusive and thus do not try to code 2 events related to the same behavioural act. For example, if main character moves close to target when hitting them, only code hitting (NOT invasion of space).
Aggressive physical	p	Prot_phs Towards Target = w Towards Bystander 1 = 1 Towards Bystander 2 = 2 Towards Bystander 3 = 3 Towards Bystander 4 = 4 Towards Bystander 5 = 5 Towards Bystander 6 = 6 Towards Bystander 7 = 7 Towards Bystander 8 = 8 Towards Bystander 9 = 9 Towards Bystander 10 = 0 ...	<p>This includes isolated hits, slaps, punches, kicks; flurries of hits, slaps, punches, kicks; pushing, shoving, grappling, headlocks, head-butts, scratches; other similar actions, ‘don’t know’.</p> <ol style="list-style-type: none"> 1. Isolated events refer to 1-2 cases of (e.g., kicks) 2. Flurry refers to 3+ instances of the same action (e.g., kicking) in quick succession. 3. Coding separate instances of the same event: a) Flurry: In the case of a flurry please code the first 2 instances only and do not code from 3+
Aggressive non-physical	o	Prot_non-phy Towards Target = w Towards Bystander 1 = 1 Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	<p>This includes aggressive isolated hand/body gestures and aggressive pointing, flurry hand gestures/aggressive pointing, isolated feinting, flurry feinting, invading space, removing clothes, running towards a person aggressively, Other / don’t know.</p> <ol style="list-style-type: none"> 1. Invading space refers to when the person stands very close to the recipient, so faces are almost touching or to when the person quickly moves into someone’s personal space. It does NOT include instances where the person is doing a physical action (such as a punch or

Event name	Code	Modifier	Rules and definitions
			<p>shove) and therefore must stand close to someone.</p> <ol style="list-style-type: none"> 2. Removing clothes includes taking off items, rolling up sleeves or throwing a coat/bag etc. to the side to free up self.
Deescalating	i	Prot_deesc Towards Target = w Towards Bystander 1 = 1 Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	<p>This includes open hand gestures, stepping/standing between people who are aggressive, pulling a person away, walking from immediate fight, light stroking, rubbing of another's body and/or face, calming hugs/embraces, reconciliatory handshakes, isolated defence, ongoing defence, Other / don't know.</p> <ol style="list-style-type: none"> 1. Walk away from immediate fight refers to when the character moves from fight but stays in the vicinity 2. Walk right away refers to when the character moves right away from the incident, such as moving across the street or leaving camera view 3. Defence: isolated refers to brief (less than 2 seconds) action whereby the character protects self (e.g., covers face) whereas ongoing defence refers to prolonged protection for 2+ seconds

Target Events

In contrast, the target is the person with whom the main protagonist's aggressive behaviours are originally oriented.

If the target performs any of the following acts towards another person, please press 'w' to start coding this target's action.

Event name	Code	Modifier	Rules and definitions
Aggressive physical	p	Tar_phs Towards Protagonist = q Towards Bystander 1 = 1 Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	Same rules and definitions as "Protagonist Events" apply
Aggressive non-physical	o	Tar_non-phy Towards Protagonist = q Towards Bystander 1 = 1 Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	
Deescalating	i	Tar_deesc Towards Protagonist = q Towards Bystander 1 = 1 Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	

Bystander Events

Bystanders are classified as all *other* people who actively contribute at least one escalatory, de-escalatory behaviour towards anyone else in the clip, or who receive at least one such behaviour.

Please start by coding the actions of bystander 1 right through (if applicable) to bystander 10. For example, if there are only three bystanders then please code bystanders 1, 2 and 3 only. If there are six bystanders then please code bystanders 1, 2, 3, 4, 5 and 6 only. The max no. of bystanders in this coding frame is 10—this can be extended if required.

If bystander 1 performs any of the following acts towards another person, please press ‘1’ to start coding this bystander 1’s action. If bystander 2 performs any of the following acts towards another person, please press ‘2’ to start coding this bystander 2’s action, etc.

Event name	Code	Modifier	Rules and definitions
Aggressive physical	p	By_phs Towards Protagonist = q Towards Target = w Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	Same rules and definitions as “Protagonist Events” apply
Aggressive non-physical	o	By_non-phy Towards Protagonist = q Towards Target = w Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	
Deescalating	i	By_deesc Towards Protagonist = q Towards Target = w Towards Bystander 2 = 2 Etc. (up to Bystander 10 = 0)	