



Department of Psychology

The Role of Social Networking in Shaping Hatred: An Exploration into User- Responses to and Influence and Permissibility of Online Hatred

SARAH ROHLFING

The thesis is submitted in partial fulfilment of the requirements for the award of the degree
of Doctor of Philosophy of the University of Portsmouth.

Date 17.01.2017

Supervised by: Professor Aldert Vrij

Dr. Stefanie Sonnenberg

Dr. Samantha Mann

Free speech includes the right “to offend, to shock or to disturb the State or any part of the population”. It does not include the right to incite violence and hatred.

Věra Jourová

General Abstract

Hate on the Internet, and particularly the role of Social Networking Sites (SNSs) in shaping online hate, the topic of investigation in this thesis, has been neglected in psychological and hate crime literature. Chapter 1 introduces the concept of online hatred (i.e. expressions of hateful content on the Internet), points out the gaps in current online hate research and emphasises the importance of studying this concept in an applied psychological context. Chapter 2 describes two studies, which together develop an explicit prejudice measure towards Roma and Travellers, as a) there was no specific one for these groups and b) they were the target groups in the study discussed within Chapter 3. The first study tested the general statement suitability of an existing prejudice measure (i.e. Levinson & Sandford's 1944 Anti-Semitism scale) to measure non-group specific prejudice. Sixteen statements were identified as suitable to measure non-group specific prejudice. The second study tested the specific appropriateness of the 16 non-group specific statements for describing stereotypes associated with Roma and Travellers. Here, 10 out of the 16 statements were rated as appropriately describing stereotypes associated Roma and Travellers. Chapter 3 describes a study that aimed to investigate the persuasive effects (i.e. changes in levels of prejudice) of online discussions on small 'like-minded' groups of participants. Groups of participants with similar levels of prejudice (i.e. low, intermediate, high) towards Roma and Travellers (the target groups) discussed the eviction of a particular British Roma and Traveller site via instant messaging online. During the discussions, a confederate expressed views which aimed at either increasing or reducing prejudice towards Roma and Travellers. Results revealed that only participants with intermediate levels of prejudice towards Roma and Travellers were influenced by the discussions (i.e. participants became more prejudiced). Yet, participants with intermediate levels of prejudice resisted online influence which aimed at reducing their levels of prejudice. Overall, the results indicate that

participants with intermediate levels of prejudice (i.e. weak attitudes) gave in to online influence, whereas those with low or high levels of prejudice (i.e. strong attitudes) resisted it. Chapter 4 explores how polarised and non-polarised YouTube users responded to racist online content. In particular, 71,000 user comments made in response to a video clip, depicting a woman on a London tram who racially abused ethnic minority passengers, were analysed using thematic analysis. The analysis revealed that the exposure to hateful online content does not lead to an automatic endorsement of hatred. More specifically, some YouTubers responded by trying to account for the woman's behaviour, as well as with hateful comments in response to viewing the hateful video clip. It also revealed that many YouTubers attempted to position themselves and other users according to their stance on racism. In addition, most responses focused on a more complex issue, namely the contestation of Britishness in relation to immigration. Chapter 5 discusses a survey, exploring the permissibility of online hatred among Social Networking Site (SNS) users. The survey specifically explored SNS users' perceptions towards what constitutes online hate, the association between online and offline hate and the role of online anonymity on expressing online hatred. Results were somewhat contradictory and thus not entirely clear. In particular, whilst most participants did not connect online with offline hatred, they blamed the victims of online hatred for their abuse but not the creators of hateful content. Further, they did not agree with legislation governing hate speech (i.e. pointing towards the permissibility of online hatred). Participants also rated hateful content to be criminal and admitted that online anonymity would aid their own anti-normative behaviour (i.e. implying online hatred is not permissible). Chapter 6 summarises the main findings from this thesis and discusses their implications, methodological considerations, and suggestions for future research.

Contents

Declaration.....	xii
List of Tables	xiii
Abbreviations and Glossary.....	xiv
Acknowledgements	xvi
Dissemination	xvii
Chapter 1 General Introduction.....	1
1.1 Overall Rationale to Research	1
1.3 Thesis Outline	2
1.4 Introduction.....	8
1.4.1 Importance of Studying Online Hate	8
1.4.2 Gaps in the Online Hate Crime Literature	12
1.4.3 Issues Concerning Online Hate Research	15
1.5 References.....	16
Chapter 2 Developing an Explicit Measure of Prejudice: Determining Attitudes Towards Roma and Travellers (Studies 1a & 1b)	27
2.1 Abstract.....	28
2.2 Introduction.....	29
2.3 Study 1a: Method.....	33
2.3.1 Ethical Considerations	33
2.3.2 Participants.....	33
2.3.3 Procedure	34
2.3.4 Materials	34
2.3.5 Data Preparation.....	36
2.4 Study 1a: Results	37
2.5 Study 1a: Discussion.....	45
2.6 Study 1b: Method.....	47
2.6.1 Ethical Considerations	47
2.6.2 Materials	47
2.6.3 Participants and Procedure.....	48
2.7 Study 1b: Results	48

2.8 Study 1b: Discussion	51
2.9 Studies 1a & 2b: General Discussion.....	51
2.9.1 Validity and Reliability Concerns.....	52
2.9.3 Conclusion	54
2.10 References.....	55

Chapter 3 The Influence of Group Discussions with ‘Like-Minded’ Others on Individuals’ Levels of Prejudice.	61
3.1 Abstract.....	62
3.2 Introduction.....	63
3.2.1 The Elaboration Likelihood Model of Persuasion	65
3.2.2. Persuasion Exerted by the Group.....	68
3.2.2.1 Group polarisation and self-categorisation.	68
3.2.2.2 Social comparison.....	69
3.2.3 Study Overview	70
3.2.4 Hypotheses.....	72
3.3 Method	72
3.3.1 Participants.....	72
3.3.2 Materials	73
3.3.2.1 Discussion topic- Dale Farm.....	73
3.3.2.2 Topic related stimulus materials.	74
3.3.2.3 Confederate training.....	74
3.3.2.4 Study access.....	75
3.3.3 Ethical Considerations	75
3.3.4 Procedure	76
3.3.4.1 Session one – determining existing levels of prejudice.....	76
3.3.4.2 Session two – the discussion.....	76
3.3.4 Data Preparation.....	80
3.4 Results.....	81
3.4.1 Pre-Task Responses and Manipulation Checks	81
3.4.1.1 Confidence to engage.....	81
3.4.1.2 Topic knowledge.....	81
3.4.1.3 Understanding of target group’s culture	82
3.4.1.4 Media influence	82

3.4.2 Post-Task Responses.....	82
3.4.2.1 Task engagement	82
3.4.2.3 Group identification following Karasawa (1991).....	83
3.4.3.4 Group commitment following Ellemers et al. (1999).....	83
3.4.2.4 Post topic knowledge	85
3.4.2.5 Changes in topic knowledge	85
3.4.2.6 Post-task understanding of the target group’s culture.....	86
3.4.2.7 Changes in perceived understanding of the target group’s culture.....	86
3.4.3 Changes in Levels of Prejudice.....	87
3.4.3.1 Participants with low levels of prejudice.....	90
3.4.3.2 Participants with intermediate levels of prejudice.....	90
3.4.3.3 Participants with high levels of prejudice.....	91
3.4.4 Extreme Cases- Large Changes in Levels of Prejudice.....	91
3.5 Discussion.....	92
3.5.1 Pro- and Counter-Attitudinal Messages.....	93
3.5.2 Attitude Strength.....	94
3.5.3 Social Influence	94
3.5.4 Expectations of Message Content.....	95
3.5.5 Personal Relevance	96
3.5.6 Experiences.....	96
3.5.7 Task Engagement.....	96
3.5.8 Limitations	97
3.5.8.1 Ecological validity and realism.....	97
3.5.8.2 Time restrictions.	98
3.5.8.3 Information processing.	98
3.5.8.4 Target group’s ‘specificness’.....	99
3.5.9 Future Directions	99
3.5.9.1 Inoculation treatments.....	99
3.5.9.1.1 Education.	99
3.5.9.1.2 Attitude strength.....	100
3.5.9.2 Online anonymity.....	100
3.5.9.2.1 Pre-task relationships.....	101
3.5.10 Conclusion	102
3.6 References.....	103

Chapter 4 “Who is really British anyway?” A Thematic Analysis of Responses to Online Hate Materials	114
4.1 Abstract.....	115
4.2 Introduction.....	116
4.3 Method	122
4.3.1 Materials	122
4.3.1.1 Data source.....	122
4.3.1.2 Ethical considerations.....	123
4.3.1.3 The YouTube video clip.....	124
4.3.1.4 Online YouTube responses.....	125
4.3.2 Analytic Strategy	125
4.3.2.1 Data preparation.....	125
4.3.2.2 Preliminary analysis.....	126
4.3.2.3 Data analysis.....	127
4.4 Results.....	128
4.4.1 Making Sense of Emma.....	131
4.4.2 Meeting Hatred with Hatred	134
4.4.3 Us versus Them.....	136
4.4.3.1 Sense of injustice.....	139
4.4.3.2. Sense of deprivation.....	141
4.4.4 Contesting Britishness	142
4.4.4.1 Britishness under threat.....	144
4.4.4.2 “We are all immigrants”.....	146
4.5 Discussion.....	148
4.5.1 Exposure to Online Hate Materials and Escalation	153
4.5.2 Limitations	155
4.5.3 Conclusion	156
4.6 References.....	158
Chapter 5 The Permissibility of Online Hatred: Exploring the Relationships between Hatefulness and Criminality, Online and Offline Hatred and the Role of Anonymity in Expressing Hatred	171
5.1 Abstract.....	172
5.2 Introduction.....	173

5.2.1 Background.....	173
5.2.2 What Constitutes Online Hate?.....	175
5.2.3 Online versus Offline Hatred.....	176
5.2.4 The Role of Online Anonymity in Expressing Online Hatred.....	177
5.2.5 Research Questions.....	178
5.3 Method.....	179
5.3.1. Ethical Considerations.....	179
5.3.2 Participants.....	179
5.3.2.1 Social networking.....	180
5.3.3 Materials and Procedure.....	180
5.3.3.1 Hatefulness and criminality.....	181
5.3.3.2 Effects of online hate on the offline world.....	182
5.3.3.3 Severity of online versus offline hate.....	183
5.3.3.4 Freedom of expression.....	184
5.3.3.5 Responsibility-accountability.....	184
5.3.3.6 Online anonymity.....	185
5.4.1 What Content Constitutes Online Hatred?.....	188
5.4.1.1 Hatefulness.....	188
5.4.1.2 Criminality.....	189
5.4.1.3 The relationship between hatefulness and criminality.....	190
5.4.2 Online versus Offline Hatred.....	190
5.4.2.1 Effects of online hate.....	190
5.4.2.2. Severity of online versus offline hatred.....	191
5.4.2.3 Freedom of expression.....	191
5.4.2.4 Responsibility/accountability.....	192
5.4.3 The Role of Online Anonymity in Expressing Hatred.....	194
5.4.3.1 Perceptions of online anonymity.....	194
5.4.3.2 Reductions in online anonymity over time.....	194
5.4.3.3 Online anonymity and subsequent behaviour.....	194
5.5 Discussion.....	196
5.5.1 What Online Content Constitutes Hatred?.....	196
5.5.2 Online versus Offline Hatred.....	197
5.5.3 The Role of Online Anonymity in Expressing Hatred.....	198
5.5.4 Limitations.....	201

5.5.6 Future Directions	202
5.5.6.1 Changing attitudes	202
5.5.6.2 Online anonymity.....	204
5.5.7 Conclusion	204
5.6 References.....	205
Chapter 6 General Discussion.....	212
6.1 General Discussion Outline	213
6.2 Overview of Main Findings.....	213
6.3 Reflexivity and Learning Gains	219
6.4 Implications of Findings	223
6.4.1 Spreading and Preventing Online Hatred	223
6.4.1.1 The Role of (Social) Identification in Shaping Hatred.....	226
6.4.1.2 The Role of Anonymity in Expressions of Online Hatred.....	228
6.4.2 Online Victimization and Rehabilitation	231
6.3. Methodological Considerations	232
6.3.1 Social Desirability.....	232
6.3.2 Ecological Validity of Experimental Study Designs	234
6.4. Directions for Future Research	235
6.5 Conclusion	236
6.6 References.....	237
Appendices.....	248
Appendix I Levinson and Sanford's (1944) Anti-Semitism (A-S) Scale	249
Appendix II Participant Consent and Instructions (Chapter 2 - Study 1a)	252
Appendix III List of 46 Adapted Attitude Statements (Chapter 2 – Study 1a).....	255
Appendix IV Participant Debrief (Chapter 2 - Study 1a)	258
Appendix V Participant Consent and Instructions (Chapter 2 – Study 1b).....	259
Appendix VI Participant Debrief (Chapter 2 – Study 1b)	261
Appendix VII Participant Consent & Instructions Session One (Chapter 3 - Study 2).262	
Appendix VIII Overall Participant Instructions Session Two (Chapter 3 - Study 2)	265
Appendix IX Fact Sheet- Dale Farm Eviction (Chapter 3 – Study 2)	267
Appendix X Manipulation Statements (Chapter 3 – Study 2)	268
Appendix XI Example Discussion Transcripts (Chapter 3 – Study 2)	271
Appendix XII Participant Debrief (Chapter 3 – Study 2).....	280

Appendix XIII Initial Coding Scheme – Content Analytic (Chapter 4 – Study 3).....	281
Appendix XIV Participant Consent and Instructions (Chapter 5 – Study 4).....	283
Appendix XV Full List of Proposed Measures’ Statements (Chapter 5).....	285
Appendix XVI Participant Debrief (Chapter 5 – Study 4)	289
Appendix XVII UPR16- Ethics Checklist	290
Appendix XVIII Favourable Ethical Opinion (Chapter 2)	291
Appendix XIX Favourable Ethical Opinion (Chapter 3).....	292
Appendix XX Favourable Ethical Opinion (Chapter 4)	293
Appendix XXI Favourable Ethical Opinion (Chapter 5).....	294
Appendix XXII “Who is Really British Anyway?” A Thematic Analysis of Responses to Online Hate Materials- Publication Version (Cyberpsychology).....	295

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Sarah Rohlfiing

WORD COUNT = 61,941 words

List of Tables

2.1 Summary table of MANOVA results illustrating suitable prejudice statements.....	38
2.2 Summary table showing the one-factor- solution ‘prejudice’ of the factor analysis (descending order).....	42
2.3 Summary table of statements’ appropriateness towards Roma and Travellers (descending order).....	49
3.1 Summary table illustrating the F and p-values of social influence covariates for the independent variables ‘level of prejudice’ and ‘type of argument’.....	89
3.2 Summary table illustrating changes in prejudice scores for both arguments (reinforcing/ refuting existing stereotypes).....	911
3.3 Summary table of extreme changes in levels of prejudice.....	922
4.1 Distribution of overarching theme and sub-theme frequencies including examples obtained from the thematic analysis.....	129
5.1 Summary table illustrating key findings concerning participants’ SNS memberships and account activity.....	1800
5.1 Summary table illustrating means, standard deviations and frequencies of participants’ hatefulness and criminality ratings in relation to their legal statuses.....	1900
5.3 Summary table of overall means, standard deviations and frequencies of participants’ (dis)agreement to the proposed attitude	

Abbreviations and Glossary

ART Scale; Anti-Roma and Traveller Scale: An explicit attitude measure developed in Chapter 2 of this thesis to determine people's levels of prejudice against Roma and Travellers (see also Chapter 3).

CMC; Computer Mediated Communication: Any human communication that occurs through the use of two or more electronic devices (e.g. chat-rooms, email, instant messaging, online forums, social networking, text messaging; McQuail, 2005)¹

DIN; Deutsches Institut für Normung (German Institute for Standardisation), which is best known in relation to paper sizes.

ELM; Elaboration Likelihood Model of Persuasion: A dual process theory which describes how attitude change might occur as a result of dealing with stimuli in one of two different routes (Petty & Cacioppo, 1986a, b).

IM; Instant Messaging: A type of online chat, which allows two or more parties to transmit information (e.g. text, images, audio files) and communicate in real-time (immediate) over the Internet.

ISIS; Islamic State of Iraq and Syria: A self-styled name of an extremist militant/radical Sunni Muslim group which rules by Wahhabi/Salafi law. Their aim is to restore an Islamic state in the region surrounding Syria, Lebanon, Israel, Jordan, the Palestinian territories, and parts of Turkey. The group is also known as Da'esh (Arabic) ISIL (Islamic State in Iraq and the Levant) and IS (Islamic State; Dictionary.com, n.d)

¹ Please note that the above sources can be found in the reference list at the end of the first Chapter.

ISP; Internet Service Provider: An organisation (commercial, community-owned, non-profit, private) that provides services for accessing/using the Internet.

ODE, Online Disinhibition Effect: Reduced or abandoned social constraints/restrictions during online communication, mediated, predominantly, by online anonymity (Suler, 2004).

SIDE; Social Identity Model of Deindividuation Effects: A social psychological model which suggests that anonymity changes the relative salience of personal versus social identity and thereby affects group behaviour (Reicher, Spears & Postmes, 1995).

SIT; Social Identity Theory: A social psychological theory which aims to explain inter-group behaviours, on the basis of a person's sense of who they are in relation to their own group membership/s (i.e. self-concept; Tajfel & Turner, 1979).

SNS; Social Networking Site: Any website that enables users to create public profiles within that website and form relationships with other users of the same site who can access their profile (Beal, n.d.).

SMP; Social Media Platform: Forms of electronic communication through which users create online communities to share information, ideas, personal messages, and other content (MiriamWebster.com, n.d.).

TA; Thematic Analysis: A common form of analysis in qualitative research, which examines and records patterns (i.e. themes) of a phenomenon closely related to a research question within and across data sets (Braun & Clarke, 2006).

TV; Television: A system for converting visual images (with sound) into electrical signals, transmitting them by radio or other means, and displaying them electronically on a screen (Oxforddictionaries.com, n.d.).

Acknowledgements

There are many people who have helped and supported me in the process of completing my thesis and I am very thankful to every one of them. First, I would like to say a massive thank you to my supervisory team, Prof. Aldert Vrij, Dr. Stefanie Sonnenberg, and Dr. Samantha Mann. You have been an excellent team, and have provided me with continuous support and encouragement throughout the completion of my PhD. I have learned so much from you all, including NEVER to design too complex studies again, to triple check EVERYTHING and how to remain patient during the recruitment process of complex studies with sensitive topics. I am very grateful for your combined expertise in a wide range of methodological approaches, as this enabled me to address at least some of the complexities online hatred presents with appropriate designs and mixed analytic methods.

Next, I would like to thank my parents for supporting me emotionally and financially during the past 5 and a bit years. Then, I would like to thank my grandmother Gisela, who was born before, and lived through, the second world war, and who shared her fascinating memories from that time with me for as long as I can remember. It was her stories which inspired me to my fascinating PhD topic.

I would also like to thank my fellow PhD students and friends, especially Zarah Vernham, Liam Satchell, Cristina Costantini, Richard Dennis, Kelvin Bateman and Sarah Kistenbrügger for their help and care whilst completing my PhD!

A special thank you goes to Toni, who has been supporting me much during the last year of completing my thesis and who continues to support me by giving me 10% of his desirable tranquillity.

Finally, I would like to thank my brilliant confederates Hannah Brock and Olly Waddup for being most motivated and reliable throughout my data collection of the group discussion (online influence) study. I could not have asked for two better people to have helped me.

Without you all, I could not have done it!

Dissemination

Conference Presentations

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2014). Online Hatred: A mixed method approach. Oral Presentation, *International Network for Hate Studies (INHS), Brighton, England, 8th-9th May 2014.*

Rohlfing, S., Näsholm, E., & Sauer, J. (2011). To look but not to see: Inattentional blindness in CCTV monitoring. Oral Presentation, *Conference of the Division of Forensic Psychology, Portsmouth, England, 22th-25th June 2011.*

Seminar Presentations

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2015). Preventing online hatred. Oral Presentation, *University of Portsmouth, Research Seminar, (guest speaker Dr. Mark Fallon) 29th May 2015.*

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2013). An Exploration of the causal relationship between the internet, prejudice and behaviour. Oral Presentation, *University of Portsmouth, Research Seminar (guest speaker Dr. Kate Houston), 26th July 2013.*

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2013). The dark side of the web? Poster Presentation, *University of Portsmouth, Postgraduate Poster Presentation Day, 4th July 2013.*

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2013). Exploring the causal relationship between the internet and hate. Oral Presentation, *University of Portsmouth, Research Seminar (guest speaker Dr. Galit Nahari), 12th February 2013.*

Rohlfing, S., Vrij, A., Sonnenberg, S., & Mann, S. (2012). An exploration of the causal relationship between the internet, prejudice and behaviour. Oral Presentation, *University of Portsmouth, Research Seminar (guest speaker Dr. Ghalit Nahari) 17th February 2012.*

Publications

Näsholm E., **Rohlfing S., & Sauer, J. D. (2014).** Pirate stealth or inattentive blindness? The effects of target relevance and sustained attention on security monitoring for experienced and naïve operators. *PLoS ONE 9(1)*. DOI: 10.1371/journal.pone.0086157

Rohlfing, S. (2014). Hate on the internet. In N. Hall, A. Corb, P. Gianassi, & J Grieve (Eds.), *International Handbook on Hate Crimes* (pp. 293-305). London: Routledge.

Rohlfing, S., & Sonnenberg, S. (2016). “Who is really British anyway?”: A thematic analysis of responses to online hate materials. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10*. DOI: 10.5817/CP2016-4-2

Chapter 1

General Introduction

1.1 Overall Rationale to Research

An awareness of hate and its consequences have been of long interest to me. I was born and grew up in Germany, where the consequences of a history concerning hate were thoroughly taught in school. My curiosity about this aspect of German history also made me question members of my family who lived in Germany during the second world war. This knowledge provided me with an initial sense of moral righteousness concerning people's diversity and their right to equality.

The specific interest in *online* hate was developed during my undergraduate degree. Here, I had the opportunity to select a hate-based unit which again reinforced my interest in the general area of hatred. During this unit, I was also given the chance to make contact with independent hate crime experts and a member of the Ministry of Justice from the area covering hate crime policy. These experts initially highlighted the immediacy and pressing importance for research into online hatred. In response, I set up an email alert to keep up to date with worldwide hate-related headlines and policy changes. Many of these alerts referred to online hatred, and particularly racial hatred expressed online. As a result, the focus shifted from a general interest into hate to a more specific online hate one. In addition, my consideration for the difficulty to construct realistic empirical work for some other forms of hatred (e.g. disability, sexual orientation) further narrowed the focus from general expressions of online those of *racial* hatred.

Moreover, the undertaking of this research programme, including user responses to online hatred, strengthened my initially developed stance on expressions of hatred. This stance is best represented by Věra Jourová's quote (see above), as it mirrors my views concerning the importance of freedom of expression whilst it sets it apart from online hate speech. In doing so, it highlights my shared view that there should be no room for online hate speech.

1.3 Thesis Outline

The central aim of this thesis is to explore the role of hateful racial online content, expressed on Social Networking Sites (SNSs) and Social Media Platforms (SMPs), in shaping people's views towards others. In line with this, it should be noted at this early stage of reading this thesis that there will be some offensive language throughout which could upset, shock or distress the reader. An additional aim of this thesis is to explore the social and social-psychological processes underlying the development of online hate. This PhD thesis comprises five studies written in the European article format, which means that the four distinctive chapters describing these studies will be in the format of a journal submission/publication. This also means that there is some repetition throughout the thesis and that each chapter will include a reference list. Both qualitative and quantitative methods will be applied. Based on a range of psychological theories and evidence, these studies explore different aspects of if and how hateful online content/views might shape other people's views and their subsequent online behaviour. All empirical work described in this thesis conforms to the British Psychological Society's (BPS) code of human research ethics (2014) and has been approved in its accordance by the University of Portsmouth Ethics Committee.

Chapter 2 Developing an Explicit Measure of Prejudice: Determining Attitudes Towards Roma and Travellers (Studies 1a & 1b).

This chapter describes two quantitative studies which together aim to develop an explicit prejudice measure towards Roma and Travellers in an online (i.e. explicit) context. The measure serves to establish changes in levels of prejudice towards these groups in the study discussed within Chapter 3. Whilst prejudice has been frequently linked to the development of hate, corresponding research into explicit expressions of negative attitudes often faces the risk of people responding in line with social desirability. There are,

however, certain groups (e.g. Roma and Travellers) which frequently elicit explicit negative prejudice because social norms do not prohibit open discrimination against them, thus making them appropriate for explicit prejudice research. Yet, no specific instrument to measure explicit prejudice towards Roma and Travellers exists. Therefore, the first study described within this Chapter uses an existing explicit prejudice measure (i.e. Levinson & Sanford's 1944 Anti-Semitism - AS scale) as a basis to develop such a new measure. Its associated stereotypes are said to overlap with the stereotypes associated with Roma and Travellers (Barnett, 2013). Based on an imagined, unspecified group, 111 participants rated 46 prejudice statements of the AS scale from which the original group affiliations had been removed to test the general suitability of the statements to measure prejudice for other, non-specific groups. A MANOVA, follow up factor analysis and reliability analysis identified 16 prejudice statements as suitable to measure prejudice towards groups other than the original target group. The second study tested the appropriateness of the 16 statements to measure prejudice towards Roma and Travellers. Seventy-four participants rated each statement for its particular relevance in describing prejudice associated with Roma and Travellers. Mean scores and reliability analysis deemed 10 of the statements appropriate to measure prejudice towards these groups. The 10 prejudice statements therefore form the new Anti-Roma and Traveller (ART) scale.

Chapter 3: The Influence of Group Discussions with 'Like-Minded' Others on Individuals' Levels of Prejudice.

This chapter describes a quantitative study, carried out across two sessions, which investigates whether online group discussions (e.g. the experimental task) could influence participants' levels of prejudice towards a specific target group (i.e. Roma and Travellers). In the first session, 124 participants completed the ART scale (developed in Chapter 2) to determine their pre-task levels of prejudice towards Roma and Travellers. It also served to

cluster participants with similar levels of prejudice into three 'like-minded' groups (low, intermediate, high prejudice) of three-to four participants, including one confederate. In the second session, the 'like-minded' groups discussed the eviction of a particular British Traveller site in an Instant Messaging (IM) chat (both explained in detail in Chapter 3). During the chat, the confederate attempted to influence participants' views by either expressing favourable or non-favourable views towards Roma and Travellers. Online influence (i.e. changes in levels of prejudice) was determined by comparing the pre- and post- task prejudice scores, measured by the ART scale. Three separate ANCOVAs (i.e. one for each prejudice group (i.e. low, intermediate, high) revealed that only participants with intermediate levels of pre-task prejudice were influenced by the online discussion. In particular, participants with intermediate levels of pre-task prejudice who were exposed to a confederate who expressed non-favourable views about Roma and Travellers during the discussion, indicated significantly higher levels of post-task prejudice. Yet, participants with intermediate levels of pre-task prejudice who were exposed to a confederate who expressed favourable views about the target groups during the discussion resisted online influence. Consequently, the evidence suggests that although most participants (i.e. those with low or high levels of pre-task prejudice) resisted online influence, those with intermediate levels of pre-discussion prejudice were most susceptible to negative online influence. Overall, these findings suggest that people's susceptibility to online influence relates to their existing levels of prejudice and the type of argument to which participants are exposed.

Chapter 4: “Who is really British anyway?” – A Thematic Analysis of Responses to Online Hate Materials

This chapter discusses a qualitative study, which explored how polarised and non-polarised Social Networking Site (SNS)/Social Media Platform (SMP) users¹ responded to one specific type of online hate material (a YouTube video clip depicting racial abuse). Seventy-one thousand YouTube comments in response to the video clip were analysed, using Thematic Analysis. Overall, the analysis revealed that the exposure to the video clip does not lead to an automatic escalation in the level of hatred expressed by YouTubers. Instead, a degree of contestation, concerning predominantly YouTubers’ national identity (i.e. here, Britishness), as well as their stance on racism, was observed. The analysis also identified four superordinate themes: (1) *Making sense of Emma*, which captures YouTubers’ efforts to explain the perpetrator’s behaviour depicted in the video clip; (2) *Meeting hatred with hatred*, which comprises YouTubers’ attempts to oppose any racist views, expressed in the video clip or by other YouTubers, with an equally hateful tone; (3) *Us versus them*, which encapsulates YouTubers’ propensities to group themselves and others, according to their stance on racism, into relevant in- and out-groups; and (4)

¹ Although Social Networking Sites (SNSs) and Social Media Platforms (SMPs) are not the same, YouTube’s commenting feature allows users to socially interact with each other (e.g. communicate/network, form relationships, share information). Social interacting is a feature attributed to both, SNSs and SMPs (see Glossary). Therefore, references made throughout this thesis, and particularly in Chapter 4, which refer to YouTube as a SNS and SMP are to be understood based on this overlap in definition but not on people’s usage or membership.

Contesting Britishness, which describes YouTubers' efforts to define and contest the meaning of Britishness amongst each other.

Chapter 5: *The Permissibility of Online Hatred: Exploring the Relationships between Hatred and Criminality, Online and Offline Hatred and the Role of Anonymity in Expressing Hatred*

This chapter describes a quantitative study, which investigates how SNS users perceive the permissibility of online hatred. One hundred and sixty-four participants took part in an online survey. The survey explored three areas of interest in relation to online hatred: (i) what constitutes online hatred, (ii) the relationship between online and offline hatred, and; (iii) the role of online anonymity in hateful expressions online. The areas were assessed across six purposely designed attitude measures, which were developed using relevant Principal Component Analyses (PCAs) and reliability analyses: (1) *Hatefulness and criminality*, in which participants rated 18 statements for their perceived hateful and criminality; (2) *effects of online hatred on the offline world*, which determined participants' attitudes towards the consequences of online hatred in offline contexts; (3) *severity of online versus offline hatred*, which explored whether online or offline hatred is perceived to be worse for its victims; (4) *freedom of expression*, which determined whether online hatred should or should not be governed by legislation; (5) *responsibility-accountability*, which explored participants' attitudes towards who should be held responsible for online hatred; and (6) *online anonymity*, which investigated how participants perceive their own online anonymity and its role in determining their subsequent behaviour. Frequency analysis, a one-way within subjects ANOVA, and follow-up paired t-tests revealed that participants largely agreed on what content was considered hateful and/or criminal and that participants' perceptions of hateful and criminality largely matched current legislation. In addition, correlations showed that

hateful content was not automatically considered criminal. A one-sample t-test between participants' overall mean scores of the *effects of online hate on the offline world* measure and the neutral score revealed that participants did not connect online with offline hatred. Moreover, means, frequencies and one-sample t-tests between participants' overall mean scores of the *severity of online versus offline hatred* measure and the neutral score showed that participants did not agree upon whether the consequences of online or offline hatred are worse for its victims. Furthermore, means, frequencies and a one-sample t-test between participants' overall mean score and the neutral score of the *freedom of expression* measure analyses revealed that more than half of the participants objected to there being legal consequences for online hatred. In addition, means, frequencies and one-sample t-tests between participants' overall mean scores for the *responsibility/accountability* measure (attributing blame for online hatred either to the victims, the posters, the police or Internet Service Providers - ISPs) and the neutral score revealed that most participants blamed the victims of online hatred for their abuse, rather than the poster, the police or ISPs. Finally, means, frequencies and one-sample t-tests between participants' overall mean scores of the *online anonymity* measure and the neutral score showed that the large majority of participants did not feel anonymous online, that participants were indifferent about whether their perceptions of online anonymity had changed over time (i.e. with the introduction of modern technology, media attention for online hate offences) and more than half declared that online anonymity determines their own (anti-normative) behaviour. Overall, these results concerning the permissibility of online hatred are mixed (i.e. some suggest its permissibility and some the opposite) and thus highlight the complexity of online hatred.

Chapter 6: *General Discussion*

This concluding chapter discusses the main findings of the thesis, whilst incorporating their practice, policy and research implications in the areas of spreading or preventing online hatred, social identification, online anonymity and victimisation and offender rehabilitation. Next, the chapter discusses methodological considerations relating to the studies within this thesis, such as social desirability and the generalisability of the findings, and concludes with a range of suggestions for future research.

1.4 Introduction

1.4.1 Importance of Studying Online Hate

Scholars generally agree that defining and conceptualising hatred is a challenging task – and its inherent complexity has given rise to a range of definitions of hate and hate crimes (Hall, 2010, 2014). Adding to this complexity are the many *forms* that expressions of hatred can take, including anger, negative stereotyping, aversion to others, negative prejudice, bullying, harassment, discrimination, bigotry, hostility, social exclusion, racism, homophobia, sectarianism, domestic violence, global/political acts of terrorism, extremist ideology or, in its most extreme form, extermination and genocide (Hall, 2014). In addition, new *means* through which hate can be expressed, such as the Internet and particularly Social Networking Sites (SNSs) and Social Media Platforms (SMPs), present further challenges for policy makers and law enforcement.

This PhD thesis examines how racial, ethnic and nationality hatred is expressed on SNSs and SMPs. It should therefore be noted that the broad term *online hatred* adopted throughout this thesis, including the findings of any empirical work reported here, does not intend to be automatically generalised to all other forms of online hatred (e.g. disability, homophobic, religious, gender, transgender hate). It should also be noted that although there is much overlap between the areas of hate, terrorism, radicalisation and extremism

references to these areas are not to be understood as being the same. Whilst arguably all of these areas contain aspects of hate and hate-motivated ideologies/views (the focus and parameters of this thesis), there are distinctive differences between them (e.g. in policies and/or motivations). For example, whilst hate crimes are generally not politically motivated and target individuals and/or groups (Green McFalls & Smith, 2003), terrorism tends to be politically motivated and tends to target large groups/nations of people (Saif-Alden Wattad, 2006). Yet, the outcomes of hate crimes and terrorism can ultimately be similar (i.e. fear and, in the worst case, the death of people). Therefore, some have argued that genocide (i.e. the most extreme form of a hate crime) is indeed a terroristic act (e.g. Cooper, 2001), whilst others have attributed hate crimes to small scale crimes (e.g. Craig, 2002). For the purpose of this thesis, potential criminal behaviours expressing online hate and hate-based crimes committed on the Internet are defined as:

Any crime or incident where the perpetrator's hostility or prejudice against an identifiable group of people is a factor in determining who is victimised. This is a broad definition. A victim does not have to be a member of the group. In fact, anyone could be a victim of a hate crime. (College of Policing, 2014, p. 3)

In recent years, hateful Internet messages and content have been increasing on SNSs, including Da'esh propaganda videos on YouTube and hateful tweets and Facebook messages (e.g. Fox News, 2016; NBC News; 2015; The Guardian, 2014). Yet, criminal convictions for inciting online hate remain overall rare, although there has been a recent global increase in arrests and convictions for online hate crimes (e.g. for inciting ethnic or religious hatred, extremism). For example, in the UK recent convictions for the incitement to (racial) hatred/violence on the Internet include Joshua Bonehill-Paine (Jewish News Online, 2015) and Anjem Choudary (Grierson, Dodd & Rodrigues, 2016). There have also been publicised cases in China, Germany, Kazakhstan, Luxembourg, Russia, and United

Arab Emirates where several people have been convicted of inciting hate on SNSs (e.g. Alkhanashvili, see: Kazakhstan Newslines, 2015; Benjamin Sch., see: Hasselmann, 2015; Mukhtar, see: Radio Free Europe Radio Liberty, 2015; Shevtsova-Valova, see: Radio Free Europe Radio Liberty, 2015; Schmitz, see: Focus Online, 2016; Sven Lau, see: Zeit Online, 2015; Tyumentsev, see: Reuters, 2015; Zhiquaing, see: Phillips, 2015).

However, the still relatively small number of convicted cases should not be read as indicating an absence of hateful online behaviour. Instead, they might be considered as presenting the tip of the iceberg. In line with this, a recent survey suggests that a quarter of UK citizens were victims of hate crimes in 2015, 28% of which were experienced online (BBC News, 2016). In line with this, the EU commission along with popular SNSs such as Facebook, Twitter, YouTube, Google and Microsoft have recently released a code of conduct to tackle and prevent online hate speech (Johari, 2016). The release therefore supports the recognition of online hate as an increasing global issue. Yet, in countries such as the UK, online hate crimes are not separated from offline hate crimes within official hate crime statistics. Consequently, there is no official record of the number of people who have experienced or are affected by online hate and thus the extent of the problem cannot be determined.

In addition, the small number of convictions might be explained by the lack of a global legal consensus of what constitutes an online hate crime and what lies within the domain of freedom of speech/expression. Specifically, in the US hate speech remains somewhat 'protected' by the First Amendment (Hall, 2005), which has in the past resulted in the inability to prosecute most cases. For example, whilst there can be prosecutions where there is a clear and immediate threat to individuals and/or groups in both online and offline contexts, what exactly constitutes a threat is harder to prove in online contexts.

In addition, there are issues relating to the location, which concerns the place where the crime was committed (e.g. posted) and to the legal position of the country in which the service is provided (e.g. the Internet server's host country). Although a hateful expression on an SNS might constitute an online hate crime in the country where it was committed, it may not be regarded as such in the country through which the service was provided. In the UK, the legal case of Sheppard and Whittle (see: Crown Prosecution Service, 2009; 5RB.com, 2010) established that offenders should be tried in the country in which the crime was committed, thus providing some clarity on the issue.

In line with this, some countries (e.g. Australia, Brazil, China, Canada, Germany, India, UAE, UK) have started to amend and introduce legislation governing online hate. Whilst the issue of an absent global legal definition for what constitutes online hate remains, these legislative changes, in line with the newly released code of conduct by the European Commission and popular SNSs, indicate first steps to deal with the problem. It also further highlights that online hate currently presents a worldwide issue and thus needs to be fully understood.

In addition to the legal issues online hate crimes present, there is a lack of understanding concerning the specific role of hateful online content in potentially influencing people's views towards others. In particular, it remains largely unknown if and how hateful online content might shape negative attitudes, prejudice and the incitement of criminal/hateful behaviour towards certain identifiable individuals and groups. As with the issue of the absence of a global definition of what constitutes online hate, there is also no global consensus on who is protected by legislation governing it.

In the UK, hate crime legislation tends to refer to those who are targeted for their perceived disability, gender-identity, race, religion or sexual orientation (Home Office, 2012), as well as those who are targeted because they are perceived to be transgender or

belong to an alternative sub-culture.² Yet illegal online hate content refers only to inciting hatred on the grounds of someone's race, religion and/or sexual orientation (True Vision, 2016). It should be noted, however, that hateful online content based on someone's disability or transgender (as well as their race, religion and/or sexual orientation) can also be considered a hate crime if it threatens or harasses individuals or groups (True Vision, 2016).

Therefore, given the large range of potential victims of online hate, an understanding of the role of hateful online content in influencing people's views is vital for efficient law enforcement and the successful outcome of criminal investigations. In particular, police require clear guidelines to be able to respond to online hate crimes adequately. Thus, a better understanding of how online hate might affect online users is necessary to implement effective preventative, counteractive and intervention strategies and to provide more effective policing of the problem.

1.4.2 Gaps in the Online Hate Crime Literature

Much of the existing literature and research on hate crime attempts to identify the factors that cause hate-motivated behaviour (Hardaker, 2010; Jacobs & Potter, 1998; Rauch & Schanz, 2013; Steinfeldt et al., 2010; Sullivan, 1999; Wojcieszak, 2010).

² Agencies of the criminal justice system (e.g. the police, Crown Prosecution Service, Prison Service and other agencies) decided in 2007 upon a common definition of five 'strands' of hate crimes to be recorded and monitored (e.g. disability, gender-identity, race, religion sexual orientation; Home Office, 2012). In 2013, Greater Manchester Police also started to record hate crimes committed against members of subcultures (e.g. Goths, Emos, The Blogging Goth, 2015) – though, they have not been adopted as an official monitored strand across the UK yet.

However, despite extensive (social) psychological research into hate, prejudice and stereotypes (e.g. Allport, 1954; Baron & Byrne, 1994; Brown, 1995), we still know little about the underlying social psychological processes that may contribute to hateful offending behaviour (Bowling, 1999; Hall, 2010b). Specifically, little is known about the extent to which hateful online content is a potential causal risk factor, that contributes to shaping negative attitudes and subsequent hateful actions (Hardaker, 2010; Rauch & Schanz, 2013; Sibbitt, 1997; Green, McFalls & Smith, 2003).

To date, research investigating (online) hate crimes has largely focused on individuals and groups who have already expressed polarised/hateful views towards others online and offline. In particular, online hate research has tended to focus either on hateful web content (e.g. Angie et al., 2011; Cammaerts, 2009; Coffey & Woolworth, 2004; Doane, 2006; Gerstenfeld, Grant & Chiang, 2003; Wojcieszak, 2010) or on the producers of hate speech (e.g. DeKoster & Houtman, 2008; Erjavec & Poler Kovačič, 2012; Ezekiel, 2002; Glaser, Dixit & Green, 2002; Green, Abelson & Garnett, 1999). Whilst this *outcome-focused* approach has been useful in identifying some specific personality traits associated with hate speech and how hateful ideologies are shared on the Internet, it is also problematic. In particular, this approach largely ignores individuals who have not yet expressed hate (are not yet polarised).

The recent introduction of the code of practise (Johari, 2016) and amendments to existing legislation and policy governing hateful online behaviour clearly suggests that governments across the world believe that SNSs, and particularly interactions with others online, can shape people's attitudes and behaviours negatively. Yet, in the context of shaping and exacerbating hate, very little is known about the specific role played by SNSs and online interactions in this context. In particular, there is a lack of empirical research

investigating the link between the exposure to online hate (i.e. online interactions, including expression of opinions, and/or hateful content) and people's (existing) attitudes.

Yet, there are some cases which support the notion that exposure to hateful online content influences or provokes hateful offline behaviour. In particular, criminal investigators have found hate-promoting videos and websites on computers seized from terrorist suspects. Specific cases include the Boston bombers (NBC News, 2015), a man who planned a terrorist attack in the style of the Lee Rigby murder (Simpson, 2016), the 2015 Charlie Hebdo attack (Spiegel Online, 2015) and the recent beheading of a girl by her Russian nanny who explicitly claimed to have been inspired to commit the act by ISIS videos (Perring, 2016).

In line with this, there is literature on the influence that films, violent games and the media have on aggression, violence and crime. This literature predominantly suggests that exposure to violent films, video games and media violence makes people more aggressive and can lead to criminal behaviour (e.g. Anderson & Bushman, 2001; 2002a; Anderson & Dill, 2000; Anderson, Murphy, in press; Bushman & Anderson, 2001; Bushman & Huesman, 2001; Dahl & Della Vigna, 2008; Funk, Bechtholdt Balacci, Pasold & Baumgardner, 2004; Johnson, Cohen, Smailes, Kasen & Brook, 2002; Sherry, 2001; Uhlman & Swanson, 2004). Given that the link between the exposure to aggressive films and aggressive behaviour has already been established, it seems reasonable to suggest that there might also be a link between the exposure to hateful online views/materials and hateful behaviour.

Moreover, whilst some have argued that social networking spaces allow for/facilitate debate, public dialogue, knowledge exchange and exposure to political difference (e.g. Brundidge, 2010; Dahlgren, 2005), others have argued that SNSs/SMPs can reinforce existing political views (i.e. echo chamber effect; Sunstein, 2001; Mutz &

Martin, 2001). The latter argument is supported by other research which found that online users have a tendency to search for, interact with and evaluate the views of ‘like-minded’ others (Adamic & Glance, 2005; Kushin & Kitchener, 2009; Stroud, 2010), as such interactions are thought to be experienced positively (selective exposure theory, Festinger, 1954). One consequence of such selective interactions is the creation of homogenous groups which share the same views. Another consequence relates to the potential for individual’s and group’s opinions to become more extreme (Sunstein, 2001), as the ever more filtered information they process prevents the exposure to other (e.g. opposing/alternative) views/arguments (Pariser, 2011).

1.4.3 Issues Concerning Online Hate Research

Online and explicit offline hate research suffer from social desirability. This is not surprising given that explicit expressions of racial attitudes might be perceived in general to be socially unacceptable (Crowne & Marlowe, 1960; Furnham, 1986; Gawronski, LeBel & Peters, 2007; Joinson, 1999; Steinfeldt et al., 2010). As a result, asking people about their attitudes towards others might result in them responding untruthfully. Whilst racial attitudes can also be measured implicitly (e.g. the implicit association test; Greenwald, McGhee & Schwartz, 1998; Payne, Cheng, Govorun & Stewart, 2005), online hate is inevitably expressed in an overt, explicit manner. Therefore, in the context of exploring the influence of hateful online views/content on people’s views, explicit measures seem to be most appropriate for online hate research.

On the basis of the evidence outlined above, and given that the (social) psychological processes underlying the development of online hate are a relatively new and under-researched area, this PhD thesis will examine *the role of hateful online content in shaping people’s attitudes towards others*.

1.5 References

- Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery*. New York, NY: ACM.
- Allport, G.W. (1954). *The nature of prejudice*. Massachusetts: Addison-Wesley Publishing Co.
- Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science, 12*, 353–359. DOI: 10.1111/1467-9280.00366
- Anderson, C. A., & Bushman, B. J. (2002a). The effects of media violence on society. *Science, 295*, 2377–2379. DOI: 10.1126/science.1070765
- Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790. DOI: 10.1037//0022-3514.78.4.772
- Anderson, C. A., & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior, 29*, 423-429. DOI: 10.1002/ab.10042
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology, 41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Barnett, R. (2013). *Jews & Gypsies: Myths & reality*. Ruth Barnett.
- Baron, R. A., & Byrne, D. (1994). *Social psychology: Understanding human interaction* (7th ed.). Massachusetts: Allyn and Bacon.

- BBC News (2016, January 27). Quarter of people have witnessed hate crime, poll suggests. *BBC News UK*. Retrieved from <http://www.bbc.com/news/uk-35415988>
- Beal, V. (n.d.). Term- social networking site. *Webopedia*. Retrieved from https://www.webopedia.com/TERM/S/social_networking_site.html
- Bowling, B. (1999). *Violent racism: Victimisation, policing and social context*. New York: Oxford University Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. DOI: 10.1191/1478088706qp063oa
- British Psychological Society (2014). *Code of human research ethics*. Retrieved from http://www.bps.org.uk/system/files/Public%20files/code_of_human_research_ethics_dec_2014_inf180_web.pdf
- Brown, R. (1995). *Prejudice: Its social psychology*. Oxford: Blackwell.
- Brundidge, J. (2010). Encountering “difference” in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks. *Journal of Communication*, 60, 680–700. DOI:10.1111/j.1460-2466.2010.01509.x.
- Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin*, 28, 1679–1686. DOI: 10.1177/014616702237649
- Bushman, B. J., & Huesmann, L. R. (2001). Effects of televised violence on aggression. In D. G. Singer, & J. L. Singer (Eds.), *Handbook of children and the media*. Thousand Oaks, CA: Sage.
- Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces: The case of north Belgian extreme right discourses. *International Journal of Cultural Studies*, 12, 555-575. DOI: 10.1177/1367877909342479

- Coffey, B., & Woolworth, S. (2004). "Destroy the scum, and then neuter their families:"
The web forum as a vehicle for community discourse? *The Social Science Journal*,
41(1), 1-14. DOI: 10.1016/j.soscij.2003.10.001
- College of Policing, (2014). *Hate crime operational guidance*. Retrieved from [http://report-
it.org.uk/files/hate_crime_operational_guidance.pdf](http://report-it.org.uk/files/hate_crime_operational_guidance.pdf)
- Cooper, H. H. A. (2001). Terrorism: the problem of definition revisited. *American
Behavioral Scientist*, 44, 881-893. DOI:10.1177/00027640121956575
- Craig, K. M. (2002). Examining hate-motivated aggression: a review of the social
psychological literature on hate crimes as a distinct form of aggression. *Aggression
and Violent Behaviour*, 7, 85-101. DOI: 10.1016/S1359-1789(00)00039-2
- Crown Prosecution Service (2009, January 8). Two guilty of inciting racial hatred against
Jews. *CPS.ORG.UK*. Retrieved from
http://www.cps.gov.uk/news/latest_news/101_09/index.html
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of
psychopathology. *Journal of Consulting Psychology*, 24, 349-354. DOI:
10.1037/h0047358
- Dahl, G., & Della Vigna, S. (2008). *Does movie violence increase violent crime?* NBER
Working Paper No. 13718.
- Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion
and deliberation. *Political Communication*, 22, 147–162.
DOI:10.1080/10584600590933160.
- DeKoster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me'.
Information, Communication & Society, 11, 1155-1176. DOI:
10.1080/13691180802266665

- Dictionary.com (n.d.). *Definition of ISIS*. Retrieved from <https://www.dictionary.com/browse/isis>
- Doane, A. (2006). What is racism? Racial discourse and racial politics. *Critical Sociology*, 32, 255-74. DOI: 10.1163/156916306777835303
- Erjavec, K., & Poler Kovačič, M. P. (2012). “You don't understand, this is a new war!” Analysis of hate speech in news web sites'. *Comments, Mass Communication and Society*, 15, 899-920. DOI: 10.1080/15205436.2011.619679
- Ezekiel, R. S. (2002). An ethnographer looks at neo-Nazi and Klan groups: The racist mind revisited. *American Behavioral Scientist*, 46, 51-71.
DOI:10.1177/0002764202046001005
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Focus Online (2016, February 22). Strafbefehl gegen Internet-Hetzer aus Bautzen. *Focus Online*. Retrieved from http://www.focus.de/regional/dresden/volksverhetzung-auf-facebook-strafbefehl-gegen-internet-hetzer-aus-bautzen_id_5305101.html
- Fox News, (2016, May 23). Muslim woman behind viral selfie hailed as heroine until hateful tweets revealed. *Fox News Europe*. Retrieved from <http://www.foxnews.com/world/2016/05/23/muslim-woman-behind-viral-selfie-hailed-as-heroine-until-hateful-tweets-revealed.html>
- Funk, J. B., Bechthold Baldacci, H., Pashold, T., & Baumgardner, J. (2004). Violence exposure in real-life, video games, television, movies and the internet: Is there desensitization? *Journal of Adolescence*, 27, 23-39. DOI: 10.1016/j.adolescence.2003.10005
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400. DOI: 10.1016/0191-8869(86)90014-0

- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193. DOI: 10.1111/j.1745-6916.2007.00036.x
- Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, 3, 29-44. DOI: 10.1111/j.1530-2415.2003.00013.x
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58, 177-193. DOI: 10.1111/1540-4560.00255
- GlobalGreen, D. P., Abelson, R. P., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice & D. T. Miller (Eds.), *Cultural divides*, (pp. 429-464). New York: Russell Sage Foundation.
- Green, D. P., McFalls, L. H., & Smith, J. K. (2003). Hate crime: An emergent research agenda. In Perry, B. (Ed.). *Hate and bias crime: A reader*. (pp. 27-48). New York, Routledge.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480. DOI: 10.1037//0033-295X.109.1.3
- Grierson, J., Dodd, V., & Rodrigues, J. (2016, August 12). Anjem Choudary convicted of supporting Islamic State. The Guardian. Retrieved from <https://www.theguardian.com/uk-news/2016/aug/16/anjem-choudary-convicted-of-supporting-islamic-state>
- Hall, N. (2005). *Hate crime*. Cullompton: Willan Publishing.

- Hall, N. (2010). Law enforcement and hate crime: Theoretical perspectives on the complexities of policing hatred. In N. Chakraborti (Ed.). *Hate crime: Concepts, policy, future directions*. (pp. 169-193). Collumpton: Willan Publishing.
- Hall, N. (2014). *Hate crime* (2nd ed.). London: Routledge.
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture*, 6, 215-242. DOI: 10.1515/jplr.2010.011
- Hasselmann, J. (2015, September 5). Polizeieinsatz nach zynischem Jubel über totes Flüchtlingskind. *Der Tagesspiegel*. Retrieved from <http://www.tagesspiegel.de/berlin/polizei-justiz/nach-rassistischer-hetze-auf-facebook-polizeieinsatz-nach-zynischem-jubel-ueber-totes-fluechtlingskind/12281654.html>
- Home Office (2012, September 13). *Hate Crimes, England and Wales 2011 to 2012*. Retrieved from <https://www.gov.uk/government/publications/hate-crimes-england-and-wales-2011-to-2012--2/hate-crimes-england-and-wales-2011-to-2012>
- Jacobs, J. B., & Potter, K. (1998). *Hate Crimes: Criminal law and identity politics*. New York: Oxford University Press.
- Jewish News Online (2015, December 17). Extremist Bonehill jailed for three years for inciting racial hatred. *Jewish News Online*. Retrieved from <http://jewishnews.timesofisrael.com/extremist-bonehill-guilty-of-inciting-racial-hatred/>
- Johari, S. (2016, June 3). EU commission & internet COS to take down online hate speech in 24 hours. *MEDIANAMA*. Retrieved from <http://www.medianama.com/2016/06/223-eu-commission-online-hate-speech/>

- Johnson, J. G., Cohen, P., Smailes, E. M., Kasen, S., & Brook, J. S. (2002). Television viewing and aggressive behavior during adolescence and adulthood. *Science*, *295*, 2468–2471. DOI: 10.1126/science.1062929
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, *31*, 433-438. DOI: 10.3758/BF03200723
- Kazakhstan Newswire (2015, August 3). New person sentenced for ‘inciting religious hatred’ in Kazakhstan. *Kazakhstan Newswire*. Retrieved from <http://www.newswire.kz/news.php?y=15&m=08&d=03#0>
- Kushin, M., & Kitchener, K. (2009). Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday*, *14*, 1–16.
- Levinson, D. J., & Sanford, R. N. (1944). A scale for the measurement of anti-Semitism. *The Journal of Psychology*, *17*, 339-370. DOI: 10.1080/00223980.1944.9917200
- McQuail, D. (2005). *McQuail’s mass communication theory* (5th ed.). London: Sage Publications.
- MiriamWebster.com (n.d.). *Definition of social media*. Retrieved from <https://www.merriam-webster.com/dictionary/social%20media>
- Mutz, D. C., & Martin, P. S. (2001). Facilitating communication across lines of political difference: The role of mass media. *American Political Science Review*, *95*, 97–114.
- NBC News (2015, March 19). Boston marathon bombing jury sees jihadist material taken from Tsarnaev devices. *NBC News Boston Bombing Trial*. Retrieved from <http://www.nbcnews.com/storyline/boston-bombing-trial/bombing-jury-sees-jihadist-material-taken-tsarnaev-devices-n326736>

- Oxforddictionaries.com (n.d.). *Definition of television*. Retrieved from <https://en.oxforddictionaries.com/definition/television>
- Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.
- Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293. DOI: 10.1037/0022-3514.89.3.277
- Perring, R. (2016, March 15). Islamic nanny who beheaded little girl was inspired by online ISIS videos. *Express*. Retrieved from <http://www.express.co.uk/news/world/652683/Burka-nanny-behead-girl-inspired-ISIS-video-Islamic-State-Islamic>
- Petty, R. E., & Cacioppo, J. T. (1986a). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.
- Petty, R. E., & Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, (pp. 123-205). New York: Academic Press.
- Phillips, T. (2015, December 22). Pu Zhiqiang given three-year suspended sentence. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2015/dec/22/pu-zhiqiang-chinese-human-rights-lawyer-sentenced-to-three-years>
- Radio Free Europe Radio Liberty (2015, July, 31). Another Kazakh sentenced for inciting ethnic hatred online. *Radio Free Europe Radio Liberty*. Retrieved from <http://www.rferl.org/a/kazakhstan-ethnic-hatred-man-sentenced/27163239.html>

- Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior*, 29, 610-615. DOI: 10.1016/j.chb.2012.11.011
- Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, 6, 161-198. DOI: 10.1080/14792779443000049.
- Reuters (2015, December, 30). Russian court jails blogger for five years for 'extremist' posts. *Reuters*. Retrieved from <http://uk.mobile.reuters.com/article/idUKKBN0UD16O20151230>
- Saif-Alden Wattad, M. (2006). Is terrorism a crime or an aggravating factor in sentencing? *Journal of International Criminal Justice*, 4, 1017-1030. DOI:10.1093/jicj/mql078
- Sherry, J. L. (2001). The effects of violent video games on aggression: A meta-analysis. *Human Communication Research*, 27, 409–431. DOI: 10.1111/j.1468-2958.2001.tb00787.x
- Sibbitt, R. (1997). *The perpetrators of racial harassment and racial violence*. Home Office Research Study No. 176. Retrieved from the Great Britain Home Office Web site: <http://www.homeoffice.gov.uk/rds/pdfs/hors176.pdf>
- Simpson, F. (2016, May 15). Man, 25, jailed over planned terrorist attack inspired by Lee Rigby murder. *News Shopper*. Retrieved from http://www.newsshopper.co.uk/news/14493976.Man__25__jailed_over_planned_terrorist_attack_inspired_by_Lee_Rigby_murder/?ref=mrb&lp=10

- Spiegel Online (2015, November 19.) 'Punching in the dark': Why Islamic state is winning the online war. *Spiegel Online International*. Retrieved from <http://www.spiegel.de/international/europe/authorities-struggle-to-control-islamic-state-online-a-1063490.html>
- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano Jr., L. A., Benton, E., & Steinfeldt M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology, 16*, 362-371. DOI: 10.1037/a0018692
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*, 556–576. DOI:10.1111/j.1460-2466.2010.01497.x.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*, 321-326. DOI: 10.1089/1094931041291295.
- Sullivan, A. (1999, September 26). What's so bad about hate? The illogic and illiberalism behind hate crime laws. *The New York Times Magazine*. Retrieved from <http://www.nytimes.com/1999/09/26/magazine/what-s-so-bad-about-hate.html>
- Sunstein, C. R. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Pacific Grove, CA: Books/Cole.
- The Blogging Goth (2015, August 18). *Sylvia Lancaster and light in Manchester*. Retrieved from <https://theblogginggoth.com/2015/08/18/sylvia-lancaster-at-darkness-and-light-in-manchester/>

- The Guardian, (2014, September 24). ISIS in duel with Twitter and YouTube to spread extremist propaganda. *The Guardian World*. Retrieved from <http://www.theguardian.com/world/2014/sep/24/isis-twitter-youtube-message-social-media-jihadi>
- True Vision (2016). *Internet hate crime: Offensive material and the law*. Retrieved from http://report-it.org.uk/reporting_internet_hate_crime
- Uhlman, E., & Swanson, J. (2002). Exposure to violent video games increases automatic aggressiveness. *Journal of Adolescence*, 27, 41-52. DOI: 10.1016/j.adolescence.2003.10.004
- Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media Society* 12, 637-655. DOI: 10.1177/1461444809342775
- Zeit Online (2015, December 15). Bundesanwaltschaft lässt Sven Lau verhaften. *Zeit Online*. Retrieved from <http://www.zeit.de/gesellschaft/zeitgeschehen/2015-12/sven-lau-verhaftung-salafisten-scharia-polizei>
- 5RB.com (2010). Judgement EWCA-CRIM-65. *5RB.com*. Retrieved from <http://www.5rb.com/wp-content/uploads/2013/10/R-v-Shepherd-and-Anor-2010-EWCA-Crim-65.pdf>

Chapter 2

Developing an Explicit Measure of Prejudice:

Determining Attitudes Towards Roma and

Travellers (Studies 1a & 1b)

2.1 Abstract

The purpose of the two studies reported in this chapter was to develop a tool in order to measure explicit prejudice towards a group that is less likely to elicit participant responses prompted by social desirability (e.g. Roma and Travellers). The measure will be applied in a subsequent study (discussed within Chapter 3) that serves to measure changes in participants' levels of prejudice in an online context. The first study used Levinson and Sanford's (1944) Anti-Semitism (AS) scale as a basis to develop this more generic measure. After removing all Jewish references from each of the original AS scale's prejudice statements, 111 participants judged 46 of the adapted prejudice statements with a non-specified group in mind. This was in order to test each statement's general suitability to measure prejudice against groups other than the original target group. A MANOVA followed by factor analysis and reliability analysis resulted in the selection of 16 non-group specific attitude statements. The second study tested these 16 prejudice statements for their specific appropriateness for Roma and Travellers. Particularly, 74 participants rated each statement for its in-/appropriateness for these groups. Mean scores and reliability analysis revealed that 10 statements could be deemed as appropriate to measure prejudice towards Roma and Travellers. The 10 statements together form the new Anti-Roma and Traveller (ART) scale, which serves to measure (changes in) levels of prejudice towards Roma and Travellers in Chapter three's study.

2.2 Introduction

In psychology, attitude research in the context of prejudice has focused on (measuring) people's implicit and explicit attitudes towards certain social groups (e.g. Black/Jewish people/women/homosexuals). Implicit attitude measures examine participants' performance during certain tasks, which, to the participant, might seem unrelated to the group in question (e.g. Greenwald, McGhee & Schwartz, 1998; Payne, Cheng, Govorun & Stewart, 2005). For example, implicit measures include recording participants' reaction times or use of word pairing words tasks. Whilst these measures do not measure attitudes directly, responses to such tasks are believed to reflect underlying attitudes (Gawronski, LeBel, & Peters, 2007).

Explicit attitude measures, on the other hand, examine people's overt responses to questions or issues and rely on the assumption that people are responding truthfully. An advantage of this approach is that explicit measures can easily be administered in the form of questionnaires or interviews. Explicit attitude measures are also appropriate for settings where researchers aim to elicit participants' free expressions of their opinions, such as in online interactions. However, here the main disadvantage of the explicit measurement approach is that participants' responses may not always be truthful. In particular, people may respond to questions in line with social desirability and therefore responses may be distorted (Jellison, McDonnell & Gabriel, 2004). This is especially relevant in contexts where the topic in question may be perceived as sensitive or where societal norms disallow the expression of particular attitudes (Crowne & Marlowe, 1960; Furnham, 1986; Gawronski et al., 2007; Joinson, 1999).

Nevertheless, there is one particular explicit attitude context that has received increased interest in recent years, namely, investigations into the expression of prejudice and hate-based attitudes online (e.g. Amichai-Hamburger, 2005; Angie et al., 2011;

Erjavec & Poler Kovačič, 2012; Joinson, 1999; McNamee, Peterson & Peña, 2010; Peddinti, Ross & Cappos, 2014; Rauch & Schanz, 2013, Steinfeldt et al., 2010; Wojcieszak, 2010). In particular, expressions of prejudice and hate-based attitudes are increasingly displayed on Social Networking Sites (SNSs), such as instant messaging, blogging, tweeting, and commenting (Citron & Norton, 2011). The emerging research interest in this area might stem from growing concerns that the accessibility, immediacy and popularity of the Internet has the potential to be exploited by those wishing to incite hatred against particular groups (Deshotels & Forsyth, 2007; Home Office, 2009).

Another reason for the interest in this area could be recent news reports of people increasingly expressing their negative attitudes online, including prejudice and racial hatred (e.g. five Scottish men, see: Hebditch, 2015; Kaiheng, see: BBC News, 2016; Konvicka, see: TOL, 2015; Nyman, see: Balona, 2015; Sakhwat Hossain, see: BD News 24, 2015; UAE national, see: Agarib, 2016;). In London alone, arrests for expressions of negative attitudes on SNSs have increased by 37% over the past five years (Corfield, 2016). Whilst not all of the cases reported in the news resulted in criminal convictions, the stories tended to stir up much debate and controversy, both online and offline. This, in turn, sometimes appeared to escalate further the prejudice and hatred initially expressed online.

In the UK, some of the cases that resulted in criminal convictions included: R v Burgess (Crown Prosecution Service, 2011), who was found guilty to inciting racial hatred via comments on Facebook; R v Hamza (Crown Prosecution Service, 2006), who was convicted of three counts of using threatening words or behaviour likely to stir up racial hatred; R v Heaton and Hannington (Crown Prosecution Service, 2010), who were found guilty of inciting racial hatred on the Aryan Strike Force's online forum; R v Hemingway (Crown Prosecution Service, 2010), who was convicted for distributing racially

inflammatory recordings on YouTube; Neil Martin (BBC News, 2006), who was charged with inciting racial hatred on Anthony Walker's condolence web page; R v Rahman, Saleem, Javed and Muhid (Crown Prosecution Service, 2007), who were found guilty of solicitation to murder and stirring up racial hatred arising out of protests over the publication of cartoons of the Prophet Muhammad; and Teenager X (Crown Prosecution Service, 2010), who pleaded guilty to inciting racial hatred for posting hate-filled videos on YouTube. It is therefore important and not surprising that researchers have become interested in this area.

The online expressions of prejudice/hate-based attitudes mentioned above are inevitably explicit, as they require people to type and subsequently respond to each other in an overt manner. Therefore, in the context of examining and evaluating online expressions of prejudice/hate-based attitudes, in the form of online interactions, explicit attitude measures seem to be the most suitable and straightforward approach. Also, as online set-ups (i.e. remote designs) do not allow for the direct monitoring of participants' actions, the use of implicit measures (e.g. reliance on reaction times) seems unsuitable.

To date, there are several explicit scales that measure prejudice towards (certain) groups, for example, the Anti-Semitism (AS) scale (Levinson & Sanford, 1944), which measures levels of prejudice against Jewish people; or the Modern Racism (MR) scale (McConahay, 1986), which measures attitudes towards Black people in America. However, when considering the nature of online hate, the groups addressed by these existing prejudice scales are said to enjoy so called 'normative protection' (Franco & Maass, 1999). The implications of 'normative protection', as with social desirability, are that people generally are less likely to give honest public responses when asked about their attitudes towards such groups, despite them potentially holding negative attitudes about them in private.

There are, however, two particular groups which continue to elicit high levels of negative prejudice and persecution within the wider society, and particularly within Europe: namely Roma and Travellers (Franco & Maas, 1999; James, 2014; Spears & Tausch, 2012). These groups are frequently referred to pejoratively as ‘Gypsies’. For the unfortunate reason that they continue to elicit much overt prejudice and their reduced level of ‘normative protection’ (Franco & Maass, 1999), Roma and Travellers present an appropriate group for investigations into explicit prejudice. Yet, to the author’s knowledge, to date, there are no explicit prejudice measures for Roma and Travellers. A prejudice measure towards Roma and Travellers could be used in a contemporary online context. Particularly, such a measure could help assess the social processes underlying the development of negative online expressions of prejudice/hate-based attitudes.

In addition, Barnett (2013) argued that the stereotypes associated with Roma and Travellers overlap with those associated with Jewish people. Therefore, it might be possible to adapt an existing prejudice measure for Jewish people, such as the AS scale, in order to create a new explicit prejudice measure for Roma and Travellers. Yet, despite the suggested overlap in stereotypes associated with both groups, it remains unclear at this stage whether it is possible to use the statements of the original AS scale to measure prejudice against other social groups. In particular, it is not clear whether participants perceive all, some, or none of the individual attitude statements that constitute the AS scale to be applicable in addressing a group or groups other than Jewish people. It is therefore important and of experimental value to first explore participants’ responses to the original AS statements, but with them addressing a group or groups other than Jewish people. Then it can be determined whether the statement content is appropriate for Roma and Travellers.

In sum, the two studies that follow aim to develop an explicit prejudice measure for online contexts. In particular, the new explicit prejudice measure serves to determine

changes in participants' levels of prejudice in a subsequent study (discussed within Chapter 3), focusing on groups which are not 'yet' normatively protected (e.g. Roma and Travellers) and therefore less likely to elicit responses which are perceived to be socially acceptable.

2.3 Study 1a: Method

2.3.1 Ethical Considerations

Participants were informed of the nature of the study (i.e. to develop a scale to measure negative attitudes towards others) and pre-warned of expressions of strong attitudes prior to their participation (see Appendix II). They were also informed that participation was anonymous insofar that participants only needed to reveal their age and gender. In addition, participants were told that the study did not intend to label participants as prejudiced against others, but instead wanted to establish that the items to be judged were scrutinised for measuring the same principle (i.e. prejudice). To minimise the risk of psychological discomfort, participants were also able to choose their own target group, which they did not reveal to anyone at any time. After their participation, participants received detailed debrief information (see Appendix IV), outlining again that the aim of the study.

2.3.2 Participants

Initially, 155 participants aged between 19 and 59 years were recruited through personal contacts. The sample consisted largely of university staff and students. There was no requirement for participants to divulge any personal, identifying details other than their age and gender. Of those, 44 participants were excluded because of their omission of multiple data points (see section 2.3.3 *Data preparation* below). This resulted in a final sample of 111 participants with the mean age of 32 years ($SD = 13$) with a gender distribution of 79 female and 32 male participants. Forty-five participants rated the

statements on paper (i.e. on university premises) and 66 online via SurveyMonkey at their homes and workplaces.

2.3.3 Procedure

The study involved participants rating 46 attitude statements for their suitability to measure prejudice towards an undisclosed, imagined, non-specified group. The statements were presented in a survey format. Participants were instructed to rate the statements in one of two ways, to which they were assigned randomly. In particular, participants were instructed to imagine and focus on a group towards which they held either negative (N= 58 participants) or positive (N= 53 participants) feelings (see Appendix II for full participant instructions). The inclusion of asking participants to rate the statements with a positively associated group in mind allowed for the judgement of whether the statements truly measured negative attitudes. They were also instructed that this group could be/but did not need to be linked by ethnicity, religion, occupation, pastime, by being a particular sports team, or anything else that linked them together and made participants either like or dislike them. Participants were also specifically instructed to only use the ‘not applicable’ option if a statement did not seem to fit their specific (imagined) target group.

2.3.4 Materials

In order to develop the new explicit prejudice measure via tapping into respondents’ prejudice to an undisclosed/imagined target group (see above), this study drew on Levinson and Sanford’s (1944) work and the attitude statements of their Anti-Semitism (AS) scale (see Appendix I for a complete list of original statements). However, before administration to participants, six out of their original 52 statements were removed because of their lack of contemporary relevance. In addition, some original statements were too specific to Jewish people and could not be easily adapted to other groups (e.g. *One thing that has hindered the Jews from establishing their own nation is the fact that*

they really have no culture of their own; instead, they tend to copy the things that are important to the native citizens of whatever country they are in; or the true Christian can never forgive the Jews for their crucifixion of Christ). The remaining 46 attitude statements were then adapted to suit a British sample (e.g. reference to White House was changed to Whitehall, to reflect the respective seats of power (see Appendix III for a list of the 46 statements).

Furthermore, all anti-Semitic references were removed and instead blank spaces were left for participants to picture their imagined target group in them. For example, participants would be presented with the following statement: *The trouble with letting _____ into a neighbourhood is that they gradually give it a typical _____ atmosphere*. As pointed out above, this was to first explore which attitude statements would, in principle, be suitable to measure prejudice towards groups other than the target group for which it was originally designed, before testing their specific appropriateness towards Roma and Travellers in a second study, and then apply them as an explicit prejudice measure for these groups in a subsequent study. For the same reason, it was necessary to include an additional 'not applicable' option. This was despite Levinson and Sanford's (1944) argument that giving participants a hypothetical neutral point (such as 'don't know' or 'not applicable') would result in the use of this option as an avoidance-strategy for undesired or disliked questions. Here, forcing participants to make an attitude judgement about a stereotype that might not only be perceived as sensitive but might also not apply to the group they have in mind could result in their refusal to make any further attitude judgements. It was therefore crucial to include a 'not applicable' option.

The response scale for each statement ranged from -3 = strong disagreement, an absolute misconception, false to +3 = firm, strong agreement, undoubtedly true in general, plus the already described 'not applicable' option. The data were transformed following

the original scoring system (-3 = 1 point, -2 = 2 points, -1 = 3 points, 1 = 5 points, 2 = 6 points and 3 = 7 points). In order to avoid inflating or deflating any measured effects, the 'not applicable' responses were scored as 0. As a result, the possible scores for each respondent in either condition ranged between 46 and 322 where low scores indicated low(er) levels or no prejudice, and high(er) scores indicated higher levels of prejudice.

2.3.5 Data Preparation

First, the data were checked for missing and anomalous values. Second, participants who responded more than five times with 'not applicable and/or skipped making prejudice judgements were excluded. This was representative of 10% of each participant's responses and thus in line with Bennett's (2001) argument that replacing more than 10% of missing data would bias statistical analyses. Third, whilst Levinson and Sanford (1944) replaced up to nine missing responses by matching them to responses of similar statements, this study replaced only up to five missing data points per participant (i.e. 'not applicable' and skipped responses) with participants' individual overall mean rating scores (Chong Guan & Saiful Bahri Yusoff, 2011), as this procedure seemed more robust. Please note that the 'not applicable' and skipped-item responses were both treated as missing data because the number of skipped items was as few as twelve total responses (0.24%) compared to 3.51% of not applicable responses throughout the remaining 111 participants' dataset.

The method of replacing missing data with participants' overall mean scores has been criticised, as it can lead to an underestimation of the standard error (Howell, 2009). Nevertheless, this only becomes a concern if either a large number of missing data points are replaced, or if the values to be replaced are not random. In this study only up to 10% of participants' individual missing data points were replaced by their own overall mean scores. Also, those values could be classified as somewhat random because under

inspection of the overall data set, the missing values were not clustered around particular items, but instead appeared randomly throughout the entire data set.

2.4 Study 1a: Results

A between-subjects MANOVA was conducted comparing the two ways in which participants had rated the prejudice statements (positively versus negatively associated group) as the only factor and the 46 prejudice statements as dependent variables. Results revealed a significant multivariate main effect for condition, Wilks' $\lambda = .30$, $F(46, 64) = 3.196$, $p < .001$, $\eta^2 = .697$. At a univariate level, 30 significant effects were revealed. However, Levene's test was significant for 13 of those 30 statements, indicating that there was no equal variance between the two attitude conditions for those items. These 13 statements were therefore excluded from further analyses. For a summary of the remaining 17 statements see Table 2.1.

Table 2.1

Summary table of MANOVA results illustrating suitable prejudice statements.

Statement	Negative		Positive		F	p	η^2
	(N=58)	(N=53)	(N=58)	(N=53)			
	M	SD	M	SD			
It would hurt the business of a large concern if it had too many _____ employees.	4.66	1.87	3.26	2.03	14.01	<.001**	.114
_____ should make sincere efforts to rid themselves of their conspicuous and irritating faults, if they really want to stop being persecuted.	4.84	2.05	3.89	2.08	15.72	<.001**	.126
War shows up the fact that the _____ are not patriotic, nor willing to make sacrifices for their country.	3.11	1.94	2.26	1.67	6.04	.016*	.053
Much resentment against _____ stems from their tendency to keep apart and to exclude people from their social life.	4.43	1.86	3.29	1.90	10.25	.002**	.086

Colleges should adopt a quota system by which they limit the number of _____ in fields which have too many of them now.	3.17	1.95	2.40	1.85	4.52	.036*	.040
_____ may have moral standards that they apply in their dealing with each other, but with others they are unscrupulous, ruthless and undependable.	3.98	2.02	2.29	2.03	7.57	.007**	.065
There is something different and strange about _____; one never knows what they are thinking or planning, not what makes them tick.	3.60	1.94	2.64	1.93	6.80	.010*	.059
The _____ problem is so general and deep that one often doubts that democratic methods can ever solve it.	4.10	2.06	2.78	1.91	12.12	<.001**	.100
The trouble with letting _____ into a neighbourhood is that they gradually give it a typical _____ atmosphere.	4.52	2.09	2.76	1.80	22.26	<.001**	.170
_____ millionaires may do a certain amount to help their own people, but little of their money goes into worthwhile national causes.	3.98	2.07	3.06	2.09	5.43	.022*	.047

_____ leaders should encourage _____ to be more inconspicuous, to keep out of professions and activities already over-crowded with _____, and to keep out of the public notice. 2.74 1.84 2.00 1.65 4.98 .028* .044

The _____ keep much to themselves, instead of taking proper interest in community problems and good government. 4.48 1.97 3.04 1.98 14.85 <.001** .120

_____ tend to remain a foreign element in society, to preserve their old social standards and to resist the western way of life. 4.37 2.08 2.62 1.83 21.97 <.001** .168

Districts containing many _____ always seem to be smelly, dirty, shabby and unattractive. 4.06 2.23 2.26 1.91 21.30 <.001** .163

There are few exceptions, but in general _____ are pretty much alike. 4.16 2.21 3.15 2.11 5.99 .016* .052

There seems to be some revolutionary streak in the _____ make-up as shown by the fact that there are so many _____ agitators. 3.45 1.95 2.65 1.99 4.77 .031* .042

_____ tend to lower the general standard of living by their willingness to do the most menial work and to live under standards that are far below average. 3.17 2.11 2.38 1.86 4.39 .038* .039

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

Responses to the remaining 17 statements were subjected to a Principal Components Analysis. The resulting Scree Plot revealed three factors with Eigenvalues above 1. Yet, the inspection of the total variance explained showed that two of the three factors accounted for only 13.10% of the total variance, leaving one factor accounting for 48.82% of the variance. Furthermore, the component matrix extracted only one factor. Sixteen out of 17 statements loaded high on factor 1 (above .50; see Table 2.2). As these statements tapped into general levels of prejudice, the factor was labelled '*Prejudice*'. The only statement (*Much resentment against _____ stems from their tendency to keep apart and to exclude people from their social life*) with a loading below .50 was subsequently excluded, leaving sixteen statements for further analysis

Table 2.2

Summary table showing the one-factor- solution 'prejudice' of the factor analysis (descending order).

Statement	Factor 1 (‘Prejudice’)	Communality
The trouble with letting _____ into a neighbourhood is that they gradually give it a typical _____ atmosphere.	.81	.69
The _____ keep much to themselves, instead of taking proper interest in community problems and good government.	.78	.67
_____ should make sincere efforts to rid themselves of their conspicuous and irritating faults, if they really want to stop being persecuted.	.77	.61
_____ millionaires may do a certain amount to help their own people, but little of their money goes into worthwhile national causes.	.76	.57
The _____ problem is so general and deep that one often doubts that democratic methods can ever solve it.	.74	.57

_____ may have moral standards that they apply in their dealing with each other, but with others they are unscrupulous, ruthless and undependable.	.73	.55
Districts containing many _____ always seem to be smelly, dirty, shabby and unattractive.	.72	.52
Colleges should adopt a quota system by which they limit the number of _____ in fields which have too many of them now.	.72	.57
There is something different and strange about _____; one never knows what they are thinking or planning, not what makes them tick.	.70	.53
_____ leaders should encourage _____ to be more inconspicuous, to keep out of professions and activities already over-crowded with _____, and to keep out of the public notice.	.69	.56
_____ tend to remain a foreign element in society, to preserve their old social standards and to resist the western way of life.	.67	.75
There are few exceptions, but in general _____ are pretty much alike.	.67	.50
It would hurt the business of a large concern if it had too many _____ employees.	.66	.44
_____ tend to lower the general standard of living by their willingness to do the most menial work and to live under standards that are far below average.	.64	.43

There seems to be some revolutionary streak in the ____ make-up as shown by the fact that there are so many ____ agitators.	.64	.41
War shows up the fact that the ____ are not patriotic, nor willing to make sacrifices for their country.	.63	.44
Much resentment against ____ stems from their tendency to keep apart and to exclude people from their social life.	.49	.72

The remaining 16 statements were subjected to a reliability analysis, revealing a Cronbach coefficient of .93. This result is in line with the high reliability of previous studies using (attitude statements of) the Anti-Semitism scale, which have ranged from .96 to .98 (Levinson & Sanford, 1944; Jones-Wiley et al., 2007).

2.5 Study 1a: Discussion

This study served as the first step towards the development of an explicit prejudice measure in order to determine changes in people's levels of prejudice in an online context in a subsequent study. Its specific purpose was to explore and identify if any prejudice statements of an existing prejudice measure (i.e. the Anti-Semitism scale) would be suitable to measure prejudice towards other groups. Results identified 16 prejudice statements as suitable to measure explicit prejudice towards a range of non-specific target groups.

With regard to reliability, one could argue that the procedure of selecting the 16 prejudice statements was more robust than the procedure of similar studies (e.g. Levinson & Sanford, 1944). For example, this study used a stricter, more conservative and less subjective scheme for replacing missing data. However, this scheme also led to the exclusion of 44 sets of participant data (28%). Despite this seemingly large exclusion rate, similar studies (e.g. Levinson & Sanford, 1944) also excluded large numbers of participants (22%) due to missing data. Here, participants' missing data might suggest that they either refused to rate some statements because they did not like them, or they might have perceived that those statements were simply 'not applicable' to the specific group they had in mind. The difference between participants' refusals or perceptions of statements as 'not applicable' cannot be determined here, as responses were treated equally as missing data. Yet, it seems unlikely that participants refused to judge the statements because they did not like them as there were only a few missing (non-judged) data points.

In the latter case, future research would need to investigate the specific target groups for which these statements are considered in-/appropriate.

In addition, the negative wording throughout the prejudice statements can be criticised in contexts where they ‘simply’ measure prejudice because it can create a general negative view in participants and thus lead to an inflation of the measured levels of prejudice. However, here the consistent negative wording of the statements should have aided the process of selecting the most suitable statements. In particular, participants should have judged the items that are most suitable to measure prejudice towards their respective imagined groups at the end of the rating scale, compared to items which are not suitable. Specifically, participants who judged the statements with a negatively associated group in mind should have indicated their strong agreement (i.e. indicating high prejudice) to them. In addition, participants who judged the statements with a positively associated group in mind should have indicated their strong disagreement (i.e. indicating no prejudice) to them. Consequently, in addition to the obvious ‘not applicable’ ratings, unsuitable statements should have received less strong participant ratings, thus discriminating between suitable and unsuitable statements.

Finally, the extent to which the 16 adapted attitude statements derived from the AS scale can be used as a suitable measure towards specific target groups (e.g. Roma and Travellers) is subject of the next study (see below). For instance, some of these statements may be perceived as accurately depicting stereotypes associated with Roma and Travellers, but others may not. In particular, the study that follows aims to test to what extent the 16 prejudice statements are appropriate to measure prejudice towards Roma and Travellers. This is, as pointed out before, to apply them in a subsequent study (see chapter 3) where they serve to determine changes in people’s levels of prejudice towards Roma and

Travellers as the target group. The next study therefore aims to determine the extent to which each statement taps into stereotypes associated with Roma and Travellers.

2.6 Study 1b: Method

2.6.1 Ethical Considerations

Participants were informed of the nature of the study (i.e. the development of a scale to measure prejudice towards Gypsies¹ for a contemporary setting) prior to participation (see Appendix V). They were also informed that the only identifiable information they needed to provide was their age and gender. In addition, participants learnt before their participation that the study intended to establish whether the statements to be judged were appropriately describing Roma and Travellers and not to measure their own levels of prejudice towards them. After participation, participants received detailed debrief information (see Appendix VI).

2.6.2 Materials

First, Roma and Travellers were specified as the target group for each of the 16 prejudice statements identified in the previous study (e.g. *The Gypsies keep much to themselves, instead of taking proper interest in community problems and good government.*). Second, response ratings were changed from dis/agreement to each statement to a 10-point Likert scale to tap the appropriateness of each statement for measuring prejudice against these particular target groups. A rating of 1 indicated that participants perceived the statement as completely inappropriate to describe, and therefore

¹ Please note that whilst this thesis refers to the target group as Roma and Travellers, the statements participants judged referred to the pejorative term ‘Gypsy’ because people tend to be more familiar with the term. However, outside of the parameter of this thesis, Roma and Travellers are seen and to be understood as two distinctive groups.

measure, prejudice towards Roma and Travellers. A rating of 10 indicated that participants perceived the statement as completely appropriate to describe, and therefore measure, prejudice towards Roma and Travellers. Only statements which received mean ratings of six and above were deemed appropriate to measure prejudice for Roma and Travellers.

2.6.3 Participants and Procedure

Seventy-four third-year undergraduate criminology students, of whom 52 were female, 21 were male and one did not indicate their gender, and with a mean age of 24 years ($SD = 10.5$) rated the 16 statements, based on their appropriateness for measuring prejudiced attitudes towards Roma and Travellers.² In particular, participants were asked to circle each corresponding number between 1 and 10 on one out of 16 Likert scales. The study was administered on paper and did not ask participants to divulge any personal, identifying details other than their age and gender.

2.7 Study 1b: Results

Descriptive statistics showed that 10 of the 16 statements received mean scores above six, indicating their suitability/appropriateness as measures of prejudice towards Roma and Travellers (see Table 2.3 below). Consequently, the six statements that received scores below six were considered inappropriate and hence excluded, leaving 10 statements for further analyses. Responses to the remaining 10 statements were then subjected to a reliability analysis, revealing a Cronbach alpha of .92. This reliability coefficient remains in line with the reliability coefficient obtained for the 16-statements in the first study (.93).

² The inspection of the data revealed that participants had rated all of the statements and therefore no data were excluded from or replaced in the analysis.

Table 2.3

Summary table of statements' appropriateness towards 'Gypsies' (descending order).

Statement	<i>M</i>	<i>SD</i>
Gypsies keep much to themselves, instead of taking proper interest in community problems and good government.	8.30*	1.61
Gypsies tend to remain a foreign element in society, to preserve their old social standards and to resist the Western way of life.	8.05*	1.92
Gypsies may have moral standards that they apply in their dealing with each other, but with others they are unscrupulous, ruthless and undependable.	7.96*	1.33
There are a few exceptions, but in general, Gypsies are pretty much alike.	7.56*	2.14
The trouble with letting Gypsies into a neighbourhood is that they gradually give it a typical Roma and Traveller atmosphere.	7.26*	2.15
Gypsy millionaires may do a certain amount to help their own people, but little of their money goes into worthwhile national causes	7.22*	2.24
Areas containing many Gypsies always seem to be smelly, dirty shabby and unattractive.	6.80*	2.73

It would hurt the business of a large concern if it had too many Gypsy employees.	6.75*	2.54
Gypsies should make sincere efforts to rid themselves of their conspicuous and irritating faults, if they really want to stop being persecuted.	6.73*	2.56
The Gypsy problem is so general and deep that one often doubts that democratic methods can ever solve it.	6.39*	2.57
Gypsies tend to lower the general standard of living by their willingness to do the most menial work and to live under standards that are far below average.	5.97	2.98
There is something different and strange about Gypsies; one never knows what they are thinking or planning, not what makes them tick.	5.59	2.84
There seems to be some revolutionary streak in the Gypsy make-up as shown by the fact that there are so many Gypsy agitators.	5.45	2.89
War shows up the fact that the Gypsies are not patriotic, nor willing to make sacrifices for their country.	5.28	2.86
Gypsy leaders should encourage Gypsies to be more inconspicuous, to keep out of professions and activities already overcrowded with them, and to keep out of the public notice.	5.07	2.54
Colleges should adopt a quota system by which they limit the number of Gypsies in fields which have too many of them now.	4.82	2.93

**Statements to be retained for further analyses and as part of the second study of this chapter.*

2.8 Study 1b: Discussion

This study was part of the development of an explicit prejudice measure towards Roma and Travellers which is to be applied in a subsequent study discussed in the third chapter. It specifically aimed to test the appropriateness and reliability of 16 prejudice statements towards these groups. Results showed that 10 out of 16 statements were rated as appropriate and reliable measures of prejudice towards Roma and Travellers. These 10 items were therefore retained and form the new explicit prejudice measure towards Roma and Travellers, which serves to determine changes in participants' levels of prejudice in a subsequent study discussed in Chapter 3.

These findings also support Barnett's (2013) suggestion that there is an overlap between the stereotypes associated with Jewish people and Roma and Travellers. Yet, the six prejudice statements participants rated inappropriate for Roma and Travellers also indicate that this overlap does not apply to all stereotypes associated with both of these groups.

2.9 Studies 1a & 2b: General Discussion

Overall, the two studies discussed above aimed to develop an explicit prejudice measure for groups that do not enjoy normative protection (e.g. Roma and Travellers) and are therefore less likely to receive responses in line with social desirability. In particular, this new prejudice measure will be applied to determine changes in people's levels of prejudice towards Roma and Travellers as the target groups in an online context in a subsequent study (discussed in detail within Chapter 3).

The two studies identified 10 prejudice statements as suitable and appropriate to measure prejudice towards Roma and Travellers. These 10 prejudice statements together form the new anti-Roma and Traveller (ART) scale. The ART scale might be seen as a

particularly well-suited tool for investigations into expressions of negative attitudes, including online hate, in overt contexts such as in online interactions.

In line with this, the length of the ART scale is also advantageous because its shortness allows for easy and fast administration. For example, the ART scale can easily be applied to determine changes in people's levels of prejudice in experimental designs, such as the one adopted in the study described in the third chapter (i.e. the study for which it was designed).

2.9.1 Validity and Reliability Concerns

Whilst the shortness of the ART makes it an easy-to-administer tool, it also limits it. Specifically, whilst the original AS scale consisted of 52 statements tapping into a broad range of content associated with the stereotyping of Jewish people, the ART scale does not cover a broad content range. In particular, the ART scale only measures some stereotypical content associated with Roma and Travellers (e.g. that they are unwilling to integrate; that the areas that they live in are always run down; that it is their own fault that they are being persecuted by society).

Consequently, the ART scale can only measure those aspects of people's prejudice towards Roma and Travellers which are addressed by its statements. The consequence is that, by not addressing the full range of content associated with the stereotyping of Roma and Travellers, the ART scale might indicate different (i.e. lower) levels of prejudice than people might hold. Yet, for the purpose of measuring some 'general' stereotypes associated with Roma and Travellers, and thus indicate a baseline measure of prejudice towards these groups, the ART scale presents a useful tool.

In addition, as pointed out previously, one could argue that the consistent negative wording of the ART scale might create a generally negative mind-set in participants. The consequence could be that participants might rate the statements and thus the group more

negatively than they actually feel. However, as Levinson and Sanford (1944) pointed out, not all of the statements are obviously anti-Semitic (here, Anti-Roma and Traveller). Specifically, they argued that participants have to understand and evaluate different levels and aspects of Anti-Semitism. Their argument is also true for the ART scale, as the statement content remains similar to that of the original scale and therefore participants would also have to understand and evaluate the particular meaning of each of the ART scale's statements.

One further point with regard to the reliability of the ART scale relates to the early developmental procedure. In particular, the removal of all Jewish references from the original Anti-Semitism scale's statements reduced the degree of consistency. Specifically, this process resulted in an inevitably larger range of groups that participants based their judgements on than the original Anti-Semitism scale. Yet, arguably, this process was necessary in order to determine if there were any suitable statements for the successful development of the ART Scale.

Finally, the gender distribution of participants in both studies was heavily balanced on females. In addition, no data on participants' nationality was obtained in either study. Moreover, the sample of the second study consisted of criminology students, which arguably were more aware and thus sensitive to the issues explored here. As a result, the results may suffer from a potential gender, nationality and expert bias and thus may not be completely representative.

Along similar lines, the participant samples were rather 'specific' because they consisted predominantly of university staff and students, a sample, one could argue to be educated and more broad-minded than the population at large. In line with this, research into the association between education and prejudice remains controversial (Kane & Kyyrö, 2001). In broad terms, there are two main assumptions about this relationship.

First, education is considered to enlighten and therefore reduce prejudice and promote social equality (Apostle, Glock, Piazza, & Suelzle, 1983; Farley, Steeh, Krysan, Jackson, & Reeves 1994; Lipset, 1960; Stouffer, 1955). Second, education can reinforce social inequality and, in doing so, foster prejudice (Jackman & Muha, 1984; Kane, 1995).

Although there are different views on the relationship between education and prejudice, the specific sample here does not seem to have influenced the results of either of the two studies. Particularly, neither of the studies measured participants' levels of prejudice per se. Instead, the first study measured the general suitability of certain prejudice statements to measure prejudice towards non-specific groups. In addition, the second study only measured whether the prejudice statements identified in the first study could appropriately measure prejudice towards specific target groups (i.e. Roma and Travellers).

2.9.3 Conclusion

The overall purpose of the two studies discussed above was to develop a new, explicit prejudice measure for the study discussed within the next chapter. There, it will determine changes in people's levels of prejudice towards particular, normatively unprotected target groups (i.e. Roma and Travellers) in an online and therefore explicit context. Measuring prejudice towards normatively unprotected groups serves to reduce the risk of eliciting socially desirable responses. The two studies identified 10 prejudice statements as suitable to measure prejudice towards Roma and Travellers, which together form the new Anti-Roma and Traveller (ART) scale. Although the ART scale does not address every negative stereotype associated with Roma and Travellers, it nevertheless presents a useful baseline measure of prejudice towards these groups.

2.10 References

- Agarib, A. (2016, January 1). Man arrested for online hate crime. *Khaleej Times*. Retrieved from <http://m.khaleejtimes.com/article/20160101/ARTICLE/301019214/1011>
- Amichai-Hamburger, Y. (2005). *The social net: Understanding human behaviour in cyberspace*. Oxford: Oxford University Press.
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology, 41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Apostle, R. A., Glock, C. Y., Piazza, T., & Suelzle, M. (1983). *The anatomy of racial attitudes*. Berkeley: University of California Press.
- Balona, P. G. (2015, March 26). Deputies: DeBary man wrote of hatred for police, harassed wife on Facebook. *The Daytona Beach News-Journal*. Retrieved from <http://www.news-journalonline.com/news/20150326/deputies-debary-man-wrote-of-hatred-for-police-harassed-wife-on-facebook>
- Barnett, R. (2013). *Jews & Gypsies: Myths & reality*. Ruth Barnett.
- BBC News (2006, October 6). Walker race hate messenger jailed. *BBC News*. Retrieved from <http://news.bbc.co.uk/1/hi/england/merseyside/5412558.stm>
- BBC News (2016, June 28). Yang Kaiheng: Singaporean jailed for hate articles on The Real Singapore site. *BBC News*. Retrieved from <http://www.bbc.co.uk/news/world-asia-36656569>
- BD News 24 (2015, June 17). Bangladesh police warn against liking or sharing posts inciting communal hatred on social media. *BDNews24.com*. Retrieved from <http://bdnews24.com/bangladesh/2015/06/17/bangladesh-police-warn-against-liking-or-sharing-posts-inciting-communal-hatred-on-social-media>

Bennett D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*. 25, 464–469. DOI: 10.1111/j.1467-842X.2001.tb00294.x

Citron, D. K., & Norton, H. L. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review*, 91, 1435-1484, University of Maryland Legal Studies Research Paper No. 2011-16.

Corfield, G. (2016, June 2). Arrests for ‘offensive’ Twitter and Facebook messages up by a third: said something naughty in London? You could be cuffed. *The Register*.

Retrieved from

http://www.theregister.co.uk/2016/06/02/social_media_arrests_up_37pc_london_section_127_communications_act/

Crown Prosecution Service (2006, February 7). Abu Hamza convicted of eleven charges.

CPS.GOV.UK. Retrieved from

http://www.cps.gov.uk/news/latest_news/105_06/index.html

Crown Prosecution Service (2007, July 18). Cartoon protestors sentenced for soliciting murder. *CPS.GOV.UK*. Retrieved from

http://www.cps.gov.uk/news/latest_news/145_07/

Crown Prosecution Service (2010, February 17). Teenager admits incitement to racial hatred over YouTube videos. *CPS.GOV.UK*. Retrieved from

http://www.cps.gov.uk/news/latest_news/teenager_admits_incitement_to_racial_hatred/index.html

Crown Prosecution Service (2010, November 15). Conviction for uploading racist videos to YouTube. *CPS.GOV.UK*. Retrieved from

http://www.cps.gov.uk/news/latest_news/140_10/index.html

- Crown Prosecution Service (2010, June 25). Michael Heaton and Trevor Hannington convicted. *CPS.GOV.UK*. Retrieved from http://www.cps.gov.uk/news/latest_news/124_10/index.html
- Crown Prosecution Service (2011). The counter-terrorism division of the Crown Prosecution Service (CPS)- cases concluded in 2011. *CPS.ORG.UK*. Retrieved from https://www.cps.gov.uk/publications/prosecution/ctd_2011.html
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. DOI: 10.1037/h0047358
- Deshotels, T. H., & Forsyth, C. J. (2007). Postmodern masculinities and the eunuch. *Deviant Behaviour, 28*, 201-218. DOI:10.1080/01639620701232961
- Erjavec, K., & Poler Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites. *Comments, Mass Communication and Society, 15*, 899-920. DOI: 10.1080/15205436.2011.619679
- Farley, R., Steeh, C., Krysan, M., Jackson, T., & Reeves, K. (1994). Stereotypes and segregation: Neighborhoods in the Detroit area. *American Journal of Sociology, 100*, 750-780. Retrieved from http://www.jstor.org/stable/2782404?seq=1#fndtn-page_scan_tab_contents
- Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology, 29*, 469-477. DOI: 10.1002/(SICI)1099-0992(199906)29:4<469::AID-EJSP938>3.0.CO;2-S
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*, 385-400. DOI: 10.1016/0191-8869(86)90014-0

- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193. DOI: 10.1111/j.1745-6916.2007.00036.x
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480. DOI: 10.1037//0033-295X.109.1.3
- Hebditch, J. (2015, November 17). Five Scots charged with religious hate after Paris attacks. *The Press and Journal*. Retrieved from <https://www.pressandjournal.co.uk/fp/news/aberdeen/753937/five-scots-charged-with-religious-hate-after-paris-attacks/>
- Home Office (2009). *The cross-government action plan*. London: Home Office.
- Howell, D. (2009). Treatment of missing data. *Statistical Home Page*. Retrieved from: www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html
- Jackman, M. R., & Muha, M. J. (1984). Education and intergroup attitudes: Moral enlightenment, superficial democratic commitment or ideological refinement? *American Sociological Review*, 49, 751-769. Retrieved from <http://www.jstor.org/stable/2095528>
- James, Z. (2014). Hate crimes against Gypsies, Travellers and Roma in Europe. In N. Hall, A. Corb, P. Giannasi & J. Grieve (Eds.), *The Routledge international handbook on hate crime* (pp. 237-248). London: Routledge.
- Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preference and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin*, 30, 629-642. DOI: 10.1177/0146167203262076

- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, *31*, 433-438. DOI: 10.3758/BF03200723
- Jones-Wiley, D. G., Restori, A. F., Lee, H. B., & Ho, M. (2007) A research note on the Levinson and Sanford anti-Semitism scale. *Perceptual and Motor Skills*, *105*, 1023-1026. DOI: 10.2466/pms.105.3.1023-1026
- Kane, E. W. (1995). Education and beliefs about gender inequality. *Social Problems* *42*, 74-90. DOI: 10.2307/3097006
- Kane, E. W., & Kyyrö, E. K. (2001). For whom does education enlighten? Race, gender, education and beliefs about social inequality. *Gender & Society* *15*, 710-733. DOI: 10.1177/089124301015005005
- Levinson, D. J., & Sanford, R. N. (1944). A scale for the measurement of anti-Semitism. *The Journal of Psychology*, *17*, 339-370. DOI: 10.1080/00223980.1944.9917200
- Lipset, S. M. (1960). *Political man*. London: Heinemann.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91-126). New York: Academic.
- McNamee, L. G., Peterson, B., L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs*, *77*, 257-280. DOI: 10.1080/03637751003758227
- Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277-293. DOI: 10.1037/0022-3514.89.3.277

- Peddinti, S., Ross, K., & Cappos, J. (2014). On the internet, nobody knows you're a dog: A twitter case study of anonymity in social networks. *Proceedings from COSN '14: The second ACM conference on Online social networks*. Retrieved from <http://cosn.acm.org/2014/files/cosn025f-peddintiA.pdf>
- Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior, 29*, 610-615. DOI: 10.1016/j.chb.2012.11.011
- Spears, R., & Tausch, N. (2012). Prejudice and intergroup relations. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An Introduction to Social Psychology* (5th ed. pp. 449-498). Chichester: Wiley.
- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano Jr., L. A., Benton, E. & Steinfeldt M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology, 16*, 362-371. DOI:10.1037/a0018692
- Stouffer, S. (1955). *Communism, conformity and civil liberties*. New York: Doubledays.
- TOL (2015, November, 25). Czech 'bloc against Islam' leader charged with hate speech. *TOL*. Retrieved from <http://www.tol.org/client/article/25292-czech-republic-islam-far-right-martin-konvicka.html>
- Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media Society 12*, 637-655. DOI: 10.1177/1461444809342775

Chapter 3

The Influence of Group Discussions with 'Like-Minded' Others on Individuals' Levels of Prejudice.

3.1 Abstract

Increasingly, people argue that hate groups use Social Networking Sites (SNSs) to influence others to promote their hateful views. Yet, to date, there is little empirical evidence on the processes underlying the development of such influence on SNSs. This study explored whether participation in online group discussions, via instant messaging chat-rooms, could influence (increase or decrease) participants' levels of prejudice. One hundred and twenty-four participants completed this two-session study. In the first session, participants completed a prejudice measure, tapping explicit prejudice towards Roma and Travellers (target group). Based on their responses, participants were divided into three groups (low, intermediate, and high level of prejudice towards these groups). In the second session, small online groups (including one confederate) of three-to-four participants with similar levels of prejudice (low, intermediate, or high) discussed the eviction of a particular British Traveller site (i.e. Dale Farm). During the discussion, the confederate used instant messaging either to try and increase or decrease participants' existing (pre-task) levels of prejudice towards Roma and Travellers. Results showed that participation in online group discussions influenced participants with intermediate levels of pre-task prejudice. In particular, participants with intermediate levels of prejudice displayed significantly higher levels of post-task prejudice towards Roma and Travellers after the confederate exposed them to messages aiming to increase their levels of prejudice. Yet, participants with intermediate levels of pre-task prejudice, who were exposed to the argument aiming to reduce their levels of prejudice, resisted influence. This suggests that whilst most participants resisted influence through online discussions, those with intermediate levels of pre-task prejudice are most susceptible to negative online influence. Therefore, people's susceptibility to online influence seems to relate to their existing levels of prejudice and the type of argument to which they are exposed.

3.2 Introduction

Most people would probably agree that the Internet, including mobile applications and other forms of digital networking, present a range of useful and positive aspects, including research, information sharing, and social networking. Some people might even find it impossible to imagine a world without the Internet. Yet, whilst it clearly offers many benefits, some researchers, policy makers, practitioners and charitable organisations are increasingly expressing their concern about people's use of SNSs to spread hateful messages and views – that is to influence others negatively (e.g. Angie et al., 2010; Association of Chief Police Officers, 2014; Foxman & Wolf, 2013; Home Office, 2009; Lee & Leets, 2002; Organisation for Security and Co-operation in Europe, 2010; Perry, 2000; Stern, 2012).

In particular, hate groups are being accused of using SNSs to influence others through means of persuasive communication strategies, including online interactions and discussions (Wojcieszak, 2010). For example, research has already demonstrated that White supremacists and other hate groups use the Internet, and increasingly SNSs, to recruit new members (Anti-Defamation League, 2005; Douglas, 2007; Gerstenfeld, Grant & Chiang, 2003) and to reinforce their views amongst 'like-minded' others (Angie et al., 2010; DeKoster & Houtman, 2009; Glaser, Dixit & Green, 2002; Green, Abelson & Garnett, 1999, Wojcieszak, 2010). Supporting this, according to Price, Nir and Cappella (2006) it is now widely accepted that group interactions can influence individuals' opinions and that individuals choose their discussion partners based on them holding similar views (Mutz, 2006). Indeed, Moscovici and Zavalloni (1969) and Myers and Bishop (1970) found that, following 'like-minded' group discussions, participants became more extreme in their existing views (i.e. more favourable/less prejudiced or less favourable/more prejudiced towards target groups).

In line with this, a recent survey revealed that more than 25% of people in the UK witnessed hate crimes in 2015, of which 28% were on SNSs (BBC News, 2016). Another survey of young people revealed that 82% had witnessed online hate on SNSs based on race, religion, gender or sexuality in 2015 (Graham, 2016). In addition, Sunstein (2001, 2009) argued that interactions between ‘like-minded’ people in online groups, including those facilitated by SNSs, can influence them insofar that they can become more extreme (i.e. polarised) in their existing views. He further argued that such influence would ultimately mobilise some people towards engaging in acts of hatred¹, a view shared by the UK Home Office (2009). Yet, whilst some have argued that physical acts of hatred tend to be carried out by the minority of those who might have been interacting with others online (e.g. Sullivan, 1999), some of these consequences (e.g. suicide bombings, nail bombs, mass shootings) nevertheless present a serious threat.

Research investigating the strategies used by hate groups in specific set-ups (i.e. on SNSs) remains limited. In particular, social influence (here, exercised during online interactions in the form of discussions between ‘like-minded’ multi-users) within the context of online hate has not yet been examined. The potential of (social/political) influence on Social Networking Sites (SNSs) and Social Media Platforms (SMPs) through means of algorithms (e.g. Wang, Cong, Song & Xie, 2010) and filters/personalising information to match/reinforce existing views (e.g. filter bubble; Pariser, 2011; echo chamber effect; Adamic & Glance, 2005; Mutz & Martin, 2001; Sunstein, 2001) has already been established in other contexts, such as political campaigns and advertising. It

¹ In this study, the majority of references made in relation to the broad term ‘hate’ refer to ethnic hate. However, it does not exclude other forms of hate, such as hate towards alternative subcultures, disability, gender, faith or transgender.

seems therefore reasonable to suggest that these individual-based concepts do also apply to social influence in online hate. Yet, to the researcher's knowledge, there are no studies on the development of hate that have assessed the influence of participation in 'like-minded' *online group discussions*.

Building on this under-researched area, this study aims to address these gaps. In particular, an aim of this study is to examine the relationship between online influence (i.e. changes in people's existing levels of prejudice) and 'like-mindedness'. Specifically, this study aims to explore if participating in online discussions with others who hold similar levels of prejudice towards a target group, can influence individuals' existing levels of prejudice towards this group. Influence is applied by a confederate in the form of statements designed to reinforce or refute existing stereotypes towards the target group. Exploring this relationship will therefore enable us to determine if online interactions do, indeed, play a role in shaping online hate (i.e. increase or decrease levels of prejudice).

First, relevant models and theoretical frameworks that explain how participants might be influenced by online interactions will be presented. This will include the elaboration-likelihood model (ELM) of persuasion (Petty & Cacioppo, 1986a, b), group polarisation (Moscovici & Zavalloni, 1969) and self-categorisation theory (Turner, Hogg, Oakes, Reicher & Wetherell, 1987). In particular, these models and frameworks will help to illustrate when and why group discussions are likely to impact on individuals' levels of prejudice. Then, a range of predictions in relation to the role played by participation in online discussions amongst multi-users (i.e. groups) and its influence on existing levels of prejudice will be outlined.

3.2.1 The Elaboration Likelihood Model of Persuasion

In order to be influential, communication - including with those who express hate - requires those targeted to decide whether they agree or disagree (i.e. will be persuaded or

not) with the positions presented to them. According to Petty and Cacioppo's (1986a, b) Elaboration Likelihood Model (ELM), the outcome of any persuasive communication attempt depends on the level of mental effort required to process the presented information, argument, or position. In particular, they suggest that there are two routes by which we can be persuaded, a central or a peripheral route. Together these mark the endpoints of a high-to-low mental effort processing continuum (Petty & Cacioppo, 1986a, b).

The central route describes people's high efforts to process the given information, whereby successful persuasion depends on their motivation and ability to process and evaluate the position or quality of the argument presented to them (Petty & Cacioppo, 1981). Here, personal relevance of the message/topic is another important factor for successful persuasion (Petty & Cacioppo, 1986b; Petty, Cacioppo & Goldman, 1981). In line with this, if we do not understand the message, we are unlikely to be motivated to process and thus be influenced by it through means of central processing (Eagly, 1974; Lee & Leets, 2002; Woodall & Burgoon, 1981).

In addition, Petty and Cacioppo (1981) argued that in order for us to change our attitudes via the central route, we need to perceive the message, at least in part, positively. Consequently, we are unlikely to change our attitudes if we perceive the argument/message negatively. In the context of online interactions shaping online hate, members of (hate) groups are likely to perceive any persuasive attempt by other 'like-minded' individuals (i.e. those who hold similar levels of prejudice towards the target group) during online interactions, as relatively positive, and thus are likely to be persuaded. Yet, when confronted with opposing views (e.g. during online interactions with 'other' out-group members on public forums and SNSs), (hate) group members are likely to perceive such views negatively and thus resist influence/change (Rauch & Schanz, 2014; Rohlfsing &

Sonnenberg, 2016). In turn, the presentation of counter-attitudinal positions might lead people to scrutinise and evaluate them in greater detail (i.e. process them centrally) and consequently form strong counter-arguments to resist them (Petty & Cacioppo, 1990). Therefore, in the context of this study, people who are exposed to counter-attitudinal arguments are not only likely to resist influence, but also might become more hateful or exacerbate their pre-existing views (i.e. polarise) through further reinforcement of their existing views (Petty & Cacioppo, 1990; Wojcieszak, 2010).

The peripheral route represents low mental effort processing, whereby successful persuasion requires little motivation and little ability to process and evaluate the position presented carefully (Petty & Cacioppo, 1981). In particular, when people lack topic knowledge or are not motivated to process the presented information, they tend to rely on cues/heuristics to help them decide whether they agree or disagree with the message content/position. Such cues include first impressions of the communicator, perceived credibility, or whether they like the communicator (Eagly & Chaiken, 1993).

Consequently, the peripheral route is more superficial than the central route and persuasive effects following peripheral processing have been found to be less enduring than those resulting from central processing (Petty & Cacioppo, 1981; 1986a). Subsequently, those who form their attitudes through central processing not only hold stronger attitudes (Krosnick, Boninger, Chuang, Berent & Carnot, 1993), but are also more resistant to counter-arguments (Petty, Haugtvedt & Smith, 2014) than those who form their attitudes through peripheral processing (Stroebe, 2012). Unsurprisingly, strong attitudes (i.e. existing ones) are considered to be more resistant to change and are held with greater certainty than weak attitudes (Ajzen & Fishbein, 2005; Fazio, 1990, O'Keefe, 2004; Petty & Krosnick, 1995).

Little empirical research has investigated strategies and processes underlying the

development of online hate. However, persuasiveness of narrative, including receptivity (e.g. Lee & Leets, 2002), and frequency of Internet use (Steinfeldt et al., 2010) have been found to increase individuals' level of prejudice and (hateful) message acceptance. Whilst this line of research largely focused on persuasion exerted by one individual on another, in the context of this study, persuasion is expected to be the result of group influence (i.e. persuasion exerted by the group or a group member on (an)other group member(s)). Therefore, next, this aspect will be considered in more detail.

3.2.2. Persuasion Exerted by the Group

In groups, people not only express their attitudes and behaviours amongst other 'like-minded' group members, they also conform to the group's norms to guide their attitudes and behaviours (Hewstone & Martin, 2012). They thereby create and maintain a 'like-minded' belief system. In group settings, including online communication between members of hate-groups, members can be influenced for many reasons. For example, members try to i) seek the group's approval and acceptance (normative influence, Deutsch & Gerard, 1955), ii) build and maintain social relationships within the group, or iii) manage their identity as a member of a 'like-minded' group (Kelman, 1958). In addition, members can be influenced because they accept the information provided by the group to be accurate (Cialdini & Trost, 1998; Hewstone & Martin, 2012; informational influence, Deutsch & Gerard, 1955). Whilst all persuasion can be incidental or deliberate, persuasion in this study will be exerted deliberately. Therefore, some additional theoretical models that explain how the group might influence individuals during online interactions will be outlined next.

3.2.2.1 Group polarisation and self-categorisation. Group polarisation suggests that group decisions become more polarised than individual ones (Cooper, Kelly & Weaver, 2003). In particular, and as discussed as part of the ELM, following interactions

with ‘like-minded’ others (e.g. group members), individuals tend to shift their views towards more extreme opinions. Whilst the ELM focuses on the cognitive routes in which information is processed in relation to persuasion, the Self-Categorisation Theory (SCT; e.g. Turner et al., 1987) focuses on the social-psychological process by which individuals define themselves as members of a particular social group (Reicher, Haslam & Hopkins, 2005; Turner, 1991). Part of this includes individuals’ internalising the group’s critical values and norms as part of their social identity. In particular, the extent to which individuals are influenced by or conform to the group depends on: (1) how much they identify with the group; (2) how consistent new ideas and proposals brought to group members are with the group’s critical (i.e. group defining) norms and values; and (3) how much group members who are trying to influence other group members are perceived as (proto)typical (i.e. are representative as group defining; Reicher et al., 2005). Specifically, these prototypical members highlight and reinforce, through persuasive arguments, the norms and similarities of one’s own in-group (i.e. ‘us’), whilst differentiating the group from ‘other’ out-groups (i.e. ‘them’; Reicher et al., 2005). This differentiating serves to maximise the differences between both groups (i.e. creating an in- and out-group division; Reicher et al., 2005; Turner, et al., 1987), which, in turn, creates the perception of extreme group norms to which non-prototypical group members, who positively identify with the group, are likely to conform (Abrams, Wetherell, Cochrane, Hogg & Turner, 1990).

3.2.2.2 Social comparison. In the context of online hate, group polarisation is not only based on maximising perceived inter-group differences, but also on the deliberate attempt to portray in-groups favourably and out-groups unfavourably (Charteris-Black, 2006, Turner et al., 1987). In particular, hateful individuals and groups tend to use the Internet, and specifically SNSs and online discussions, to recruit new members and share/reinforce their ideology amongst ‘like-minded’ people, which, in turn, provides them

with a positive and salient social identity (Angie et al., 2011, Beirich, 2014; DeKoster & Houtman, 2008). For example, studies investigating the persuasive power of online interactions in the context of promoting and exacerbating hate (e.g. Bostdorff, 2004; Douglas, McGarty, Bliuc & Lala, 2005; McNamee, Petersen & Peña, 2010; Rauch & Schanz, 2014) have found that hate groups use strategies, such as the reinforcement of existing negative stereotypes (Steinfeldt, et al., 2010), the presentation of biased knowledge (Bostdorff, 2004), ‘out-grouping’ through creations of identity threats (Douglas, et al., 2005) and the reinforcement of a shared group identity (McNamee et al., 2010). In line with the self-categorisation account, these strategies not only maximise in/out-group differences (i.e. categorical thinking), and thereby create an even stronger sense of ‘like-mindedness’ amongst in-group members, they also have the ability to polarise the views of hate group members towards out-groups, which serves to create inter-group conflict (Gerstenfeld et al., 2003; Glaser, et al., 2002; Green et al., 1999). This inter-group conflict can then lead to hostility towards members of these out-groups via in-group mobilisation (Angie et al., 2011; Klein, Spears & Reicher, 2007), and can act as the justification for hateful actions towards out-group members (Blazak, 2001; Bostdorff, 2004; Douglas et al., 2005).

3.2.3 Study Overview

A 2 (type of argument: pro versus counter-attitudinal) x 2 (levels of prejudice: pre-/post-task level of prejudice) factorial study will investigate the persuasive effects of online discussions (i.e. changes in levels of prejudice, whereby the discussion serves as the experimental task) on groups of ‘like-minded’ participants (here, participants with low, intermediate or high levels of prejudice towards a specific target group).

The discussion topic will relate to the eviction of a particular Roma and Traveller site, namely Dale Farm, whereby Roma and Travellers represent the target group.² This specific group was selected based on the sensitive nature of the task (i.e. being exposed to and potentially expressing prejudice towards Roma and Travellers) and their unfortunate fate to continue eliciting high levels of explicit negative prejudice and persecution within the wider society, and particularly within Europe (Franco & Maas, 1999; James, 2014; Spears & Tausch, 2012). Therefore, expressions of prejudice against Roma and Travellers should be considered normative by some or all of the participants.

The dependent variable, changes in levels of prejudice towards Roma and Travellers, is measured through a self-report attitude measure (i.e. the ART Scale – see Chapter 2) on six-point Likert scales. The two independent variables are the type of argument and levels of prejudice. The type of argument is operationalised by two positions, namely whether the argument is in line with (i.e. pro) or counter-attitudinal to (i.e. anti) its recipient's views. Levels of prejudice are operationalised by two points of measurement (i.e. before and after the discussion task), hence this study will be carried out in two sessions. 'Like-mindedness' is operationalised by participants' similar levels of existing (i.e. pre-task) prejudice, which were categorised as low, intermediate and high. Group identification, in terms of its influence on participants' views (Karasawa, 1991) and

² For the purpose of this study Roma and Travellers have been grouped together because peoples' perceptions of these groups are often interchangeable. This is also evidenced by the pejorative term 'Gypsies', which is often used to refer to these groups. Yet, it should be noted that outside of the parameter of this study, Roma and Travellers are two distinctive groups.

commitment to the group (Ellemers, Kortekaas & Ouwerkerk, 1999) will be considered as additional mediating variables to online influence.

3.2.4 Hypotheses

Based on previous persuasion and online hate research, it is predicted that ‘like-minded’ participants with intermediate levels of prejudice towards Roma and Travellers (i.e. weaker attitudes) will be influenced in the direction of the message source (Ajzen & Fishbein, 2005; Fazio, 1990; Lee & Leets, 2002; O’Keefe, 2004; Petty & Krosnick, 2014), whereas participants with low and high levels of prejudice towards Roma and Travellers (i.e. strong attitudes) will resist online influence (Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Petty & Cacioppo, 1981; Sunstein, 2001; Wojcieszak, 2010; **Hypothesis 1**). It is also anticipated that, in combination with the type of argument (i.e. pro-/counter-attitudinal), social influence, in terms of influence exerted by other group members and group commitment, will further facilitate online influence for those with intermediate levels of prejudice towards Roma and Travellers (Abrams et al., 1990; Reicher et al., 2005; **Hypothesis 2**).

3.3 Method

3.3.1 Participants

Initially, 231 participants (aged between 17 and 66, $M = 25.5$, $SD = 10.7$) were recruited from the University student and staff population, as well as from the general public. A total of 107 participants were excluded before completing the second session (34 did not respond to the invitation for the second session, 21 did not show for the second session, six withdrew after the first session, and 46 fell into a condition which had already been filled), leaving 124 participants who completed both sessions. Participants were allocated to one of six conditions, depending on their existing prejudice levels towards the target group (low, intermediate or high prejudice; see method section- session one below)

and their random allocation to one of two within-subjects conditions (i.e. messages aiming at either increasing or reducing prejudice towards the target group). This resulted in the allocation of 20 participants with low levels of prejudice to the condition which aimed at reducing their levels of prejudice; another 21 participants with low levels of prejudice to the condition which aimed at increasing their levels of prejudice; 22 participants with intermediate levels of prejudice to the condition which was aimed at reducing their level of prejudice; another 21 participants with intermediate levels of prejudice to the condition which was aimed at increasing their levels of prejudice; 20 participants with high levels of prejudice to the condition which was aimed at reducing their levels of prejudice; and a final 20 participants with high levels of prejudice to the condition which was aimed at increasing their levels of prejudice. There were no significant differences between the six groups in terms of gender ($p = .160$) or age ($p = .591$).

3.3.2 Materials

3.3.2.1 Discussion topic- Dale Farm. As already mentioned, the discussion topic related to the eviction of a particular Traveller site, called Dale Farm. Dale Farm used to be the largest Traveller site in the UK, which, up until October 2011, housed more than 1,000 residents, who were predominantly Irish Travellers and some Roma. The six-acre plot of land is located within the green belt, a governmental policy for controlling urban growth, in the Basildon District in Essex. The site used to be a scrap and breaker's yard, which was bought by members of the Travelling community. Half of Dale Farm was legal with planning permission for occupation as a Traveller site, whereas the other half was established without authorisation and planning permission. In October 2011, after 10 years of legal disputes costing the council an estimated £18 million pounds, the High Court ruled the eviction of Dale Farm's residents, which led to the forcible eviction of its residents.

3.3.2.2 Topic related stimulus materials. Out of online newspaper articles concerning the Dale Farm eviction, 11 specific pro- and 12 anti-eviction statements were developed, also bearing in mind the stereotypes of the Anti-Roma and Traveller (ART) scale (see Appendix X for the full list of statements). In particular, these statements consisted of favourably or unfavourably worded positions about the eviction and they were aimed at either reinforcing or refuting existing negative stereotypes and thus prejudice towards Roma and Travellers. Pro Dale Farm Eviction (i.e. reinforcing prejudice) statements included for example, *I cannot believe that Basildon Council took so long to evict those Gypsies and that their best explanation to be allowed to evict them was that the Gypsies did not have planning permission!; The reputation of being thieves and uncivilised must be true, otherwise why does this reputation follow them literally everywhere they go?!; or I would not want to live next to or in the same area as Gypsies!* Anti Dale Farm Eviction (i.e. refuting prejudice) statements included: *I cannot believe that Basildon Council evicted the Gypsies just because they didn't have planning permission!; It is even the Gypsies' land, they bought it off the scrap yard owner and so should they be allowed to do what they like?!; or I think it was also the old stereotypes of Gypsies being thieves and the general misunderstanding of the general public towards Gypsy culture, putting pressure on the council to get rid of them.*

3.3.2.3 Confederate training. Two confederates were trained to introduce statements about Roma and Travellers, during the task (i.e. the discussion) without raising suspicion amongst the participants. First, the confederates received materials (i.e. newspaper articles, advocate web sites) which contained general messages aiming at either reinforcing or refuting negative stereotypes and thus prejudice towards the target group. These served to familiarise themselves with overall stereotypes and prejudice associated with Roma and Travellers. Next, confederates received the specific pro- and counter-Dale

Farm eviction statements, which they used to exert online influence. Confederates then practised their role during approximately five trial chat discussions, during and after which they received feedback from the researcher.

3.3.2.4 Study access. Participants completed the entire study online and in two sessions. Each session required participants to complete a range of task related questions (see sessions one and two below), which were administered via the online research tool ‘SurveyMonkey’. Participants accessed each questionnaire by clicking on relevant web links, provided in emails.

3.3.3 Ethical Considerations

This study conforms to the British Psychological Society’s (BPS) ethics guidelines on Internet-Mediated Research (IMR, 2013). For example, it sought participants’ consent prior to their participation, ensured their anonymity and maintained a level of control over the discussion task. Yet, the true aims of the study had to be withheld from participants until after their participation because of the risk of responses being influenced by social desirability. In particular, such knowledge could have acted as form of inoculation treatment and could thus have prevented any influence. Therefore, pre-task instructions only informed participants that the study explored online interactions and that it would involve two sessions including a live online chat with other participants (for full instructions, see Appendices VII & VIII). After participation, participants received detailed information about the study, including the exact purpose and manipulations of the study and the role of the confederate (see Appendix XII). This was vital to avoid any risk of ‘creating’ prejudiced or even racist participants (i.e. permanently/long-term increase their levels of prejudice towards Roma and Travellers). Participants were also given the opportunity to send feedback to the researcher which was taken up by some participants.

Such feedback was positive and encouraging throughout and stated their newly gained understanding of aspects of online hatred and their agreement to the nature of the research.

3.3.4 Procedure

3.3.4.1 Session one – determining existing levels of prejudice. Participants completed the ART scale (Cronbach's alpha = .90) via an online link to the survey software SurveyMonkey. The scale was developed specifically for this study (discussed in detail in Chapter 2). The scores of this measure determined participants' existing levels of prejudice towards Roma and Travellers which, in turn, allowed allocating participants with similar prejudice scores (i.e. low, intermediate, high) into groups of two or three for the second session.

Participants also completed some filler statements, which aimed to disguise the target statements (i.e. those concerning to Roma and Travellers and, at the same time, served to prevent participants from guessing the full aim of the study. The content of the filler statements was loosely based on the explicit prejudice statements towards Roma and Travellers, but instead addressed bankers and benefit cheats (e.g. *Bankers' behaviour is so entrenched that no democratic actions can ever change it; A major fault of the bankers is their conceit and arrogance. They believe that they are untouchable; On the whole, benefit cheats have probably contributed less to Western life than anyone else; or The trouble with letting benefit cheats into a nice area is that they gradually drag it down and give it a bad name*). The statements were presented in random order, which further served as a disguise for the target group.

3.3.4.2 Session two – the discussion. A unique chatroom on the website 'Chatzy' was created for each group prior to the start of their discussion. Two or three participants plus one confederate received their instructions explaining when/how to enter the chatroom and necessary web links to the study and the chatroom, approximately 15 minutes before

their participation. This time frame allowed enough time for participants to familiarise themselves with the requirements and procedure, but at the same time, reduced the risk for participants to carefully research the topic (i.e. be influenced) before the beginning of the task.

Next, participants followed the web link to the study, where they were presented with a short unbiased³ summary of the Dale Farm Eviction (see Appendix VIX). The summary served to increase participants' processing ability of the argument, as it provided them with a range of facts related to the discussion topic. Also, given that the evidence on the persuasive communication surrounding the effectiveness of stories and statistics remains inconclusive (e.g. Burgoon, 1990; Krauss & Chiu, 1998), the summary consisted of both a narrative-like structure and statistical elements. This was to appeal to more participants.

Then participants rated three, six-point scale questions relating to the extent of their perceived pre-task topic knowledge, the extent to which they understood the target group's culture before completing the task and their anticipated confidence to engage in the task (i.e. *I feel I have good knowledge and understanding of the recent eviction of the Travellers at Dale Farm; I have a good understanding of the Gypsy and Traveller culture and their ways of living; I am confident that I will fully engage with the others during the online*

³ The creation of the summary involved an initial collection of relevant newspaper articles from which known facts were extracted which were then collated to a summary. The calibration/piloting process was predominantly based on a pragmatic, face-validity approach and not tested formally. However, both, professional stake holders (i.e. an academic expert and relevant member of the Ministry of Justice) and other academics judged the summary to be 'unbiased'.

interactions). The scales ranged from 1- strongly disagree to 6- strongly agree. Low scores indicated little perceived topic knowledge/understanding of the target group's culture and little confidence to engage in the task and vice versa. Whilst, the questions concerning participants' perceived pre-task topic knowledge and understanding of the target group's culture provided pre-task information concerning the target group, the anticipated confidence to engage in the task served to motivate and thus increase participation. Participants were then prompted, via a message on the screen, to postpone the completion of any other questions until they completed the task (i.e. the discussion) and meanwhile enter the chatroom by opening it in another browser window via a second web link. Participants only used their individually assigned participant numbers to identify themselves to each other to reduce the risk of responses being given in line with social desirability. The researcher monitored participants' chatroom attendance and notified the confederate (via mobile phone) when to join the chatroom in order to keep the confederate's disguise. The confederate was only instructed to join the chatroom if at least two participants had already joined it because the groups required at least two participants.

Once all participants had entered the chatroom, the researcher instructed them to begin discussing their views on the Dale Farm Eviction. The confederate then tended to start the conversation by asking the other participants for their views (see Appendix XI for two examples of the discussion groups). Depending on participants' responses, s/he then immediately started to introduce the 11 or 12 specific pro or anti Dale Farm eviction statements. There was no particular order in which the statements had to be introduced, nor constraint to use all statements every time. Instead, the statement-choice depended on participants' responses to each statement and to each other. This was to reduce the risk of the discussion appearing scripted, to maintain the confederate's disguise as another participant to the other participants, and to allow for a level of continuation in the

conversation flow. It was also the confederate's responsibility to keep the discussion on topic.

There were 49 discussion groups in total- each lasted approximately 30 minutes. The end of the discussion was predominantly determined by conversation flow and by the confederate's ability to introduce as many of the specific pro or anti Dale Farm eviction statements as possible. In particular, once the conversation flow slowed down (i.e. response times became notably slower/delayed), it became repetitive, or all appropriate statements had been discussed, the confederate informed the researcher (via mobile phone) who, in turn, instructed the participants to leave the chatroom and return to the remaining questions in the first browser window.

These questions concerned participants' post-task topic knowledge and their post-task understanding of the target group's culture. The questions served to determine pre-to post-task changes in perceived task knowledge and understanding of the target group's culture and were thus also measured on six-point scales (i.e. *I now have a good understanding of the Gypsy culture and their ways of living*). Participants then completed the ART scale again (Cronbach's alpha = .90), which allowed for the comparison between participants' pre- and post-task prejudice scores and thus determine changes in their levels of prejudice.

In addition, to determine whether people were influenced by other group members, and whether this (social) influence would, in turn, mediate online influence, participants completed one of Karasawa's (1991) identification with group members' subscale items (i.e. *Are there many members of the group who have influenced your thoughts and behaviour?*). Moreover, in order to determine to what extent participants felt committed to their discussion group and, in turn, whether group commitment mediates online influence, participants completed Ellemers et al.'s (1999) three-item group commitment subscale (i.e.

I would like to continue working with my group, I dislike being a member of my group and I would rather belong to a different group). In particular, Karasawa's (1991) group identification item, and Ellemers et al.'s (1999) group commitment items (Cronbach's $\alpha = .67$) were used as co-variates to determine whether social influence, in the form of group identification and group commitment, would mediate online influence (i.e. changes in participants' levels of prejudice). These statements were rated on 4-point response scales and subsequently ranged from 1 = no identification with/commitment to the group to 4 = strong identification with/commitment to the group. The items *I dislike being a member of my group* and *I would rather belong to a different group* were reverse scored.

Finally, participants completed two questions concerning whether the media influenced their views towards Roma/Travellers between sessions (i.e. *Since completing the first part of this study, have you seen any media reports concerning Gypsies/Travellers?* and *If you have seen any media reports concerning Gypsies/Travellers since you completed the first part of this study, do you think that this has influenced your opinion towards Gypsies/Travellers?*). This was to determine the extent to which any changes in participants' levels of prejudice were attributed to media reports instead of the online discussion.

3.3.4 Data Preparation

All explicit prejudice statements were rated on 6-point response scales with no neutral point (i.e. no '0') and subsequently ranged from -3 = strong disagreement, to +3 = strong agreement. The data were then transformed, following the procedure explained in Chapter 2, so that -3 = 1 point, -2 = 2 points, -1 = 3 points, 1 = 5 points, 2 = 6 points, 3 = 7 points. As a result, the possible scores for the target group ranged from 10 to 70 (i.e. total score range of 60), where low scores indicated low levels of prejudice and high scores indicated high levels of prejudice. During the first session, the total score range was

divided by three in order to balance the scores for the three possible prejudice groups (low, intermediate, high). As a result, and given that possible scores ranged from 10 to 70, participants with scores ranging between 10 and 30 were categorised as having low levels of prejudice, participants with scores ranging from 31 to 50 were categorised as endorsing intermediate levels of prejudice, and those scoring from 51 to 70 were categorised as demonstrating high levels of prejudice.

3.4 Results

3.4.1 Pre-Task Responses and Manipulation Checks

3.4.1.1 Confidence to engage. The majority of participants were confident about their engagement during the discussion. In particular, 87 % of participants scored four or higher on the six-point Likert scale; whereby 1 indicated low and 6 indicated high confidence to engage in the task. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) being the between subjects factor and participants' perceived confidence to engage in the task being the dependent variable. The ANOVA revealed that there was no significant difference in participants' confidence to engage in the task between the three prejudice groups, $F(2, 121) = 1.869, p = .159, \eta_p^2 = .030$. These results suggest that participants in the low, intermediate and high prejudice groups were equally confident to engage in the task.

3.4.1.2 Topic knowledge. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' perceived pre-task topic knowledge as the dependent variable. The ANOVA revealed that there were no significant differences in participants' perceived topic knowledge, $F(2, 121) = .851, p = .429, \eta_p^2 = .014$. These results indicate that participants in all three prejudice groups had approximately the same amount of perceived topic knowledge before participating in the task.

3.4.1.3 Understanding of target group's culture. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' perceived pre-task understanding of the target group's culture as the dependent variable. The ANOVA revealed that there were no significant differences in participants' perceived understanding of the target group's culture, $F(2, 121) = 2.291, p = .106, \eta^2 = .037$. These results indicate that participants across all three prejudice groups (i.e. low, intermediate, high) had approximately the same amount of perceived understanding of the target group's culture before they completed the task.

3.4.1.4 Media influence. With regard to news/media influence, 96% of participants reported not to have seen any media reports concerning the target groups between the two study sessions. In addition, 3% of the sample reported to have seen media reports but that these did not influence their views concerning the target group. Only one participant reported that s/he had seen media reports concerning the target group since the beginning of the study which had influenced his/her opinions towards the target group.

3.4.2 Post-Task Responses

3.4.2.1 Task engagement. The majority of participants indicated that they felt engaged during the discussion, with 91.1% of the sample scoring 4 or higher on the 6-point Likert scale. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' perceived task engagement as the dependent variable. The ANOVA revealed a significant difference in participants' perceived task engagement, $F(2, 121) = 3.304, p = .040, \eta^2 = .052$. Tukey post-hoc tests revealed that participants with high levels of prejudice perceived themselves as significantly more engaged in the discussion ($M = 5.35, SD = .89, 95\% \text{ CI } [5.02, 5.68]$) compared to participants with low levels of prejudice ($M = 4.76, SD$

= 1.34, 95% CI [4.43, 5.08]), $p = .032$. However, there was no significant difference in perceived task engagement between participants with low and intermediate levels of prejudice ($M = 5.12$, $SD = .85$, 95% CI [4.80, 5.43]), $p = .261$ nor between participants with high and intermediate levels of prejudice ($p = .569$).

3.4.2.3 Group identification following Karasawa (1991). A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' perceived influence from other group members' ratings as the dependent variable. The ANOVA revealed that participants' perceived influence from other group members' ratings differed significantly across the three prejudice groups, $F(2,121) = 4.191$, $p = .004$, $\eta^2 = .086$. Tukey post-hoc tests revealed that participants with low levels of pre-task prejudice indicated significantly less perceived influence from (an)other group member(s) ($M = 1.81$, $SD = .13$, 95% CI [1.54, 2.07]) compared to participants with intermediate levels of pre-task prejudice ($M = 2.37$, $SD = .13$, 95% CI [2.11, 2.63]), $p = .009$. In addition, participants with intermediate levels of pre-task prejudice indicated significantly higher levels of perceived influence from (an)other group member(s) than participants with high levels of pre-task prejudice ($M = 1.85$, $SD = .14$, 95% CI [1.58, 2.12]), $p = .018$. However, there was no significant difference in perceived influence ratings between participants with low and high levels of pre-task prejudice), $p = .970$.

3.4.3.4 Group commitment following Ellemers et al. (1999). A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and the three group commitment item ratings as the dependent variables. The ANOVA revealed that there was no significant difference between participants' perceived group commitment, in terms of them wanting to

continue with the group, across the three prejudice groups, $F(2, 121) = .053, p = .948, \eta^2 = .001$.

In addition, it revealed that participants' perceived group commitment, in terms of their dislike of other group members, differed across the three prejudice groups, $F(2, 121) = 20.635, p < .001, \eta^2 = .253$. Tukey post-hoc tests revealed that participants with low levels of pre-task prejudice indicated significantly less dislike for (a) member(s) of their group ($M = 3.12, SD = .157, 95\% \text{ CI } [2.81, 3.43]$) than participants with intermediate levels of pre-task prejudice ($M = 3.65, SD = .153, 95\% \text{ CI } [3.35, 3.95]$), $p = .045$, but indicated significantly more dislike for (a) member(s) of their group than participants with high levels of pre-task prejudice ($M = 2.25, SD = .159, 95\% \text{ CI } [1.94, 2.56]$), $p < .001$. Furthermore, participants with intermediate levels of pre-task prejudice indicated significantly more dislike for (a) member(s) of the discussion group than participants with high levels of pre-task prejudice ($p < .001$).

The ANOVA also revealed that participants' group commitment, in terms of their desire to belong to a different group, differed across the three prejudice groups, $F(2, 121) = 24.794, p < .001, \eta^2 = .291$. Tukey post-hoc tests revealed that participants with low levels of pre-task prejudice indicated significantly less desire to belong to a different group ($M = 3.05, SD = .147, 95\% \text{ CI } [2.76, 3.34]$) than participants with intermediate levels of prejudice ($M = 3.63, SD = .144, 95\% \text{ CI } [3.34, 3.91]$), $p = .016$ but indicated significantly more desire to belong to a different group than participants with high levels of pre-task prejudice ($M = 2.18, SD = .149, 95\% \text{ CI } [1.88, 2.47]$), $p < .001$. Furthermore, participants with intermediate levels of pre-task prejudice indicated significantly more desire to belong to a different discussion group than participants with high levels of pre-task prejudice ($p < .001$). Taken together, these results show that participants with intermediate levels of pre-task prejudice reported the least group commitment, since they rated the highest levels of

dislike of (an)other group member(s) and the greatest desire to belong to a different discussion group. In contrast, participants with high levels of pre-task prejudice reported the most group commitment, since they rated the least dislike of (an)other group member(s) and the least desire to belong to a different discussion group.

3.4.2.4 Post topic knowledge. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' post-task topic knowledge ratings as the dependent variable. The ANOVA revealed that participants' perceived post-task knowledge ratings differed amongst the three prejudice groups, $F(2,121) = 5.231, p = .007, \eta^2 = .080$. Tukey post-hoc tests revealed that participants with low levels of prejudice reported significantly less post-task topic knowledge ($M = 4.00, SD = 1.16, 95\% \text{ CI } [3.63, 4.37]$) compared to participants with high levels of prejudice ($M = 4.65, SD = .98, 95\% \text{ CI } [4.34, 4.96]$), $p = .003$. Participants with low levels of prejudice also reported significantly less post-task topic knowledge than participants with intermediate levels of prejudice ($M = 4.53, SD = .74, 95\% \text{ CI } [4.31, 4.76]$), $p = .013$. There were no significant differences in post-task topic knowledge ratings between participants with high and intermediate levels of prejudice ($p = .590$).

3.4.2.5 Changes in topic knowledge. A 3 (level of prejudice: low, intermediate, high) x 2 (level of topic knowledge: pre-/post-task) mixed-design ANOVA was conducted, with level of prejudice (low vs. intermediate vs. high) as the between-subjects factor, level of topic knowledge (pre-/post-task topic knowledge) as the within-subjects factor, and changes in participants' perceived topic knowledge ratings as the dependent variable. There was no significant main effect for the level of prejudice on participants' perceived topic knowledge, $F(2, 121) = 2.478, p = .088, \eta^2 = .039$. The main effect of level of topic knowledge was significant, Wilks $\lambda = .418, F(1, 121) = 169.239^b, p < .001, \eta^2 = .582$.

Results showed that participants' perceived topic knowledge increased from pre- ($M = 2.58$, $SD = .135$, 95% CI [2.31, 2.85]) to post-task ($M = 4.40$, $SD = .087$, 95% CI [4.22, 4.57]). Finally, there was no significant interaction effect between participants' levels of prejudice and their levels of topic knowledge, Wilks $\lambda = .975$, $F(2, 121) = 1.574^b$, $p = .211$, $\eta^2 = .025$. Overall, these results suggest that although participants' perceived topic knowledge increased after they participated in the task, these increases did not differ between the prejudice groups.

3.4.2.6 Post-task understanding of the target group's culture. A one-way between-subjects ANOVA was conducted with level of pre-task prejudice (low vs. intermediate vs. high) as the between subjects factor and participants' understanding of the target group's culture ratings as the dependent variable. The ANOVA revealed that participants' understanding of the target group's culture ratings differed between the three prejudice groups, $F(2, 121) = 6.084$, $p = .003$. Tukey post-hoc test showed that participants with low levels of prejudice reported significantly less post-task understanding of the target group's culture ($M = 3.59$, $SD = 1.40$, 95% CI [3.14, 4.03]) than participants with high levels of prejudice ($M = 4.50$, $SD = .96$, 95% CI [4.19, 4.81]), $p = .001$. There were no differences in post-task understanding of the target group's culture between participants with low and intermediate levels of prejudice ($M = 4.07$, $SD = 1.14$, 95% CI [3.72, 4.42], $p = .063$) or those with intermediate and high levels of prejudice ($p = 1.00$).

3.4.2.7 Changes in perceived understanding of the target group's culture. A 3 (level of prejudice: low, intermediate, high) x 2 (level of understanding of the target group's culture: pre-/post-task) mixed-design ANOVA was conducted with level of prejudice (low vs. intermediate vs. high) as the between-subjects factor, level of understanding of the target group's culture (pre-/post-task) as the within-subjects factor, and changes in levels of participants' understanding of the target group's culture as the

dependent variable. There was a significant main effect for the level of prejudice on participants' perceived understanding of the target group's culture, $F(2, 121) = 4.714, p = .011, \eta^2 = .072$. Tukey post-hoc tests revealed that participants with low levels of prejudice rated their understanding of the target group's culture after participating in the task significantly lower ($M = 3.45, SD = .169, 95\% \text{ CI } [3.15, 3.82]$) compared to participants with high levels of prejudice ($M = 4.23, SD = .171, 95\% \text{ CI } [3.89, 4.56]$), $p = .008$. In addition, there was no significant difference in changes of perceived understanding of the target group's culture between participants with low and intermediate levels of prejudice ($M = 3.81, SD = .165, 95\% \text{ CI } [3.49, 4.14]$), $p = .510$ after participating in the task. Furthermore, it showed that participants with intermediate levels of prejudice were not significantly different in their understanding of the target group's culture ratings compared to participants with high levels of prejudice ($p = .259$) after they participated in the task. The main effect for participants' levels of understanding of the target group's culture was statistically significant Wilks $\lambda = .850, F(1, 121) = 21.373^b, p < .001, \eta^2 = .150$, indicating that participants' understanding for the target group's culture scores increased from before ($M = 3.63, SD = .108, 95\% \text{ CI } [3.42, 3.85]$) to after they participated in the discussion group ($M = 4.05, SD = .106, 95\% \text{ CI } [3.84, 4.26]$). Finally, there was no significant level of prejudice X level of understanding of the target groups' culture interaction, Wilks $\lambda = .975, F(2, 121) = 1.529^b, p = .211, \eta^2 = .025$. These results suggest that although participants' perceived understanding of the target group's culture increased from pre- to post-task, the significant differences in these increases were only observed between participants with low and high levels of prejudice.

3.4.3 Changes in Levels of Prejudice

Three 2 (levels of prejudice: pre-/ post-task) x 2 (type of argument: pro-/anti Roma/Traveller) mixed-design ANCOVAs, one for each prejudice group (low,

intermediate, high), were conducted with levels of prejudice (pre-/post-task) as the within-subjects factor and type of argument (pro- vs. anti-Roma/Traveller) as the between-subjects factor, and changes in participants' prejudice ratings as the dependent variable. In addition, group identification, in terms of influence exerted by other group members (i.e. Karasawa, 1991), and group commitment (i.e. Ellemers et al., 1999) was controlled for as covariates. None of the covariates had an effect on the independent variables (see Table 3.1), thereby not supporting Hypothesis 2. For clarity, the results of the three mixed-design ANCOVAs, presented next, are broken down into the three prejudice groups (i.e. low, intermediate and high levels of prejudice).

Table 3.1

Summary table illustrating the *F* and *p*-values of social influence covariates for the independent variables 'level of prejudice' and 'type of argument'.

Level of Prejudice	Independent Variable	Karasawa (1991)			Ellemers et al. (1999)								
		Influence by other group members			Want to continue with Group			Dislike of members of the group			Desire to belong to a different group		
		<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Low	Level of Prejudice (pre-/post-task)	.163	.689	.005	.903	.348	.025	.079	.781	.002	.044	.835	.001
	Type of Argument	.009	.925	.000	3.491	.070	.091	.151	.700	.004	.795	.379	.022
Intermediate	Level of Prejudice (pre-/post-task)	.012	.913	.000	.865	.358	.023	.083	.774	.002	.513	.478	.014
	Type of Argument	.643	.428	.017	.009	.925	.000	.002	.969	.000	.496	.486	.013
High	Level of Prejudice (pre-/post-task)	.036	.850	.000	.072	.791	.002	1.131	.295	.032	2.008	.166	.056
	Type of Argument	.545	.465	.016	.048	.828	.001	.039	.845	.001	.413	.525	.012

**p* < .05

3.4.3.1 Participants with low levels of prejudice. As expected, there was no significant main effect for level of prejudice, $F(1, 35) = 1.805, p = .188, \eta_p^2 = .049$, indicating that participants' levels of prejudice did not change from pre- to post-task. Also, as expected, there was no significant main effect for the type of argument on participants' prejudice ratings, $F(1, 35) = 1.490, p = .230, \eta_p^2 = .041$, indicating that there was no difference in the persuasiveness of the two types of arguments and thus changes in participants' levels of prejudice, which, in turn, supports Hypothesis 1. Finally, there was no significant interaction effect between level of prejudice and type of argument for participants with low levels of prejudice, $F(1, 35) = .231, p = .634, \eta_p^2 = .007$. These results further suggest that participants with low levels of pre-task prejudice towards Roma and Travellers were overall resistant to online influence.

3.4.3.2 Participants with intermediate levels of prejudice. There was no significant main effect for level of prejudice, $F(1, 37) = 1.572, p = .218, \eta_p^2 = .041$, suggesting that participants' levels of prejudice did not change from over the course of the study. Also, there was no significant main effect for the type of argument on participants' prejudice ratings, $F(1, 37) = 1.481, p = .231, \eta_p^2 = .038$, suggesting that the type of argument did not influence participants' levels of prejudice. There was a significant level of prejudice X type of argument interaction, $F(1, 37) = 7.197, p = .011, \eta_p^2 = .161$. The means and standard deviations for this interaction are shown in Table 3.2. A simple main effects analysis demonstrated that participants with intermediate levels of pre-task prejudice were significantly more prejudiced after being exposed to the argument aimed at increasing their levels of prejudice towards Roma and Travellers, Wilks $\lambda = .810, F(1, 39) = 9.173^a, p = .004, \eta_p^2 = .190$. Yet, participants with intermediate levels of pre-task prejudice resisted online influence aimed at reducing their levels of prejudice towards Roma and Travellers, Wilks $\lambda = .987, F(1, 39) = .510^a, p = .480, \eta_p^2 = .013$.

Table 3.2

Summary table illustrating changes in prejudice scores for both arguments (reinforcing/refuting existing stereotypes).

Time/Session	Task Manipulation				
	Reinforcing Existing			Refuting Existing	
	Stereotypes			Stereotypes	
	M	SD	n	M	SD
Pre-Task	41.19	5.03	22	41.55	5.29
Post-Task	45.86	8.38	22	40.55	9.24

3.4.3.3 Participants with high levels of prejudice. There was no significant main effect for level of prejudice on participants' prejudice ratings, $F(1, 34) = 2.546, p = .120, \eta_p^2 = .070$, indicating that participants' levels of prejudice did not change in the course of the study. Also, as expected, there was no significant main effect for the type of argument on participants' prejudice ratings, $F(1, 34) = .021, p = .886, \eta_p^2 = .001$, indicating that the type of argument (pro- or counter-attitudinal) did not determine changes in participants' levels of prejudice. This further supports Hypotheses 1. No significant level of prejudice X type of argument interaction was found for participants with high levels of prejudice, $F(1, 34) = .003, p = .956, \eta_p^2 = .000$. Overall, these results suggest that participants with high levels of pre-task prejudice towards Roma and Travellers were resistant to online influence.

3.4.4 Extreme Cases- Large Changes in Levels of Prejudice.

As previously pointed out, Sullivan (1999) suggested that the number of people who go on to commit acts of hatred tend to be a small minority. Therefore, it was decided to explore the number of extreme cases within the current sample (i.e. those who changed

their levels of pre-task prejudice across entire categories, which is equivalent to 20 points or more). Four participants of the 124 (3.2%) were identified (see Table 3.3).

Table 3.3

Summary table of extreme changes in levels of prejudice.

Level of Prejudice (pre-task)	Level of Prejudice (post-task)	Number of Points Changed	Exposure/Condition
Low	Intermediate (almost high)	21	Increasing Prejudice
Intermediate	High	21	Increasing Prejudice
High	Intermediate (almost low)	24	Reducing Prejudice
High	Low	27	Reducing Prejudice

3.5 Discussion

The aim of this study was to explore whether online interactions, in the form of live discussions between groups of ‘like-minded’ participants (i.e. those holding low, intermediate, or high levels of prejudice towards specific target groups) would influence participants’ existing levels of prejudice towards this group.

Overall, most participants indicated increased topic knowledge and understanding of the target group’s culture after their participation in the online discussion. In addition, online influence, in terms of changes in levels of prejudice, affected participants with intermediate levels of pre-task (i.e. existing) prejudice who were exposed to messages which were aimed at reinforcing stereotypes and thus increasing prejudice. Moreover, and as expected overall, participants in all other groups resisted attempts to influence them online. In particular, participants who resisted online influence attempts included those with low and high levels of pre-task prejudice who were exposed to messages which were

aimed at either refuting or reinforcing stereotypes and thus decrease or increase prejudice. In addition, participants with intermediate levels of pre-task prejudice who were exposed to messages that were aimed at refuting stereotypes and thus decrease prejudice also resisted attempts to influence them online.

3.5.1 Pro- and Counter-Attitudinal Messages

The finding that participants with either low or high levels of prejudice resisted influence attempts supports Hypothesis 1. In particular, participants with low levels of pre-task prejudice are likely to have perceived the messages which were aimed at decreasing prejudice towards the target group as pro-attitudinal and thus supportive of their existing views. Likewise, participants with high levels of pre-task prejudice would have perceived messages that aimed at increasing prejudice as pro-attitudinal and thus supportive of their existing views. As such, participants with either low or high levels of prejudice who were exposed to pro-attitudinal messages should have had no reason to change their attitudes towards the target group. This finding is also in line with previous research (e.g. Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Petty & Cacioppo, 1981; Sunstein, 2001; Wojcieszak, 2010).

Additionally, counter-attitudinal messages which were aimed at increasing prejudice towards the target for participants with low pre-task prejudice, and discussions which aimed at decreasing prejudice for participants with high levels of pre-task prejudice are most likely to have been perceived negatively as they opposed participants' existing views. As such, participants with either low or high levels of pre-task prejudice should have resisted counter-attitudinal positions. This finding is supported by existing research (e.g. Petty & Cacioppo, 1981, 1990; Wojcieszak, 2010).

3.5.2 Attitude Strength

One could argue that participants' low and high levels of pre-task prejudice scores indicate that they hold strong attitudes towards the target groups, whereas participants with intermediate levels of pre-task prejudice do not. In line with existing attitude strength research (i.e. Ajzen & Fishbein, 2005; Fazio, 1990; O'Keefe, 2004; Petty & Krosnick, 2014), strong attitudes were therefore predicted to be resistant to change whereas weak attitudes were not (i.e. supportive of Hypothesis 1). In particular, participants with weaker attitudes towards the target group (i.e. intermediate levels of pre-task prejudice) were persuaded by online influence which aimed at increasing their levels of pre-task prejudice. Yet, participants with intermediate levels of prejudice (i.e. those with weaker attitudes) resisted online influence which was aimed at reducing prejudice. These findings therefore only partially support Hypothesis 1. In addition, these findings support Lee and Leet's (2002) argument that online hate groups would benefit from neutral audiences as the persuasive effect tends to be strongest for such audiences.

3.5.3 Social Influence

With regard to perceived influence by other group members and group commitment, here too participants with intermediate levels of pre-task prejudice indicated the biggest differences compared to participants with low and high levels of pre-task prejudice. In particular, participants with intermediate levels of pre-task prejudice indicated the highest levels of perceived influence by other group members. In addition, participants in this group indicated the highest levels of dislike of other members and the greatest desire to belong to a different discussion group compared to participants with low and high levels of pre-task prejudice. As such, although participants with intermediate levels of prejudice indicated the highest levels of influence by other group members, they also identified the least with the group. In line with this, the results of this study suggest that social influence,

in terms of influence exerted by other group members and participants' group commitment, did not facilitate online influence regardless of the type of argument (i.e. pro- counter-attitudinal), thereby not supporting Hypothesis 2.

One could argue that these findings, concerning participants with intermediate levels of prejudice, hint at a level of uncertainty about their perceived identity as group members. In fact, the concept of categorising participants as holding intermediate levels of prejudice towards the target groups, suggests less defined group norms and values in comparison to groups holding low or high levels of prejudice. In particular, for participants with intermediate levels of prejudice towards the target groups, the views expressed by the confederate during the discussion were neither fully in agreement nor disagreement with their existing attitudes compared to participants with low or high levels of prejudice towards the target group. As a result, the confederate's views could have resulted in increased uncertainty about, and thus decreased identification with, the group for participants with intermediate levels of pre-task prejudice (Abrams et al., 1990, Hogg, 2000; Turner, 1985).

3.5.4 Expectations of Message Content

The finding that participants with intermediate levels of prejudice towards the target group resisted online influence aimed at reducing prejudice suggests that the existing (negative) stereotypes about this culture might be relatively resistant to influence. One reason for this resistance might lie in participants' expectations about this culture. In particular, Parrott (1995) argued that unexpected message content, including discrepant cultural messages, can capture individuals' attention to a message's position (e.g. central processing) and thus result in their resistance to it. Here, the presentation of Roma and Travellers (i.e. the target group) in a way that aimed to refute the usual stereotypes might have resulted in greater scrutiny of the message content (i.e. central processing) and

consequently led participants to resist social influence, which ran contrary to their existing views.

3.5.5 Personal Relevance

Alternatively, overall resistance to online influence might be rooted in specific aspects of the discussion. In particular, one aspect of all discussions concerned the large amount of money the eviction of the Traveller site (i.e. a financial issue related to the target group) had cost the council. Here, it was observed that participants with all levels of prejudice (i.e. low, intermediate and high) discussed this aspect at great length compared to other aspects of the discussion. Given the financial recession during the time of data collection, it is possible that participants might have identified with the situational financial strain, making the cost of the eviction personally relevant to them, which, in turn, might have served as a particularly persuasive cue (Petty & Cacioppo, 1986b; Petty, Cacioppo & Goldman, 1981).

3.5.6 Experiences

Moreover, it was observed during several discussion groups that personal experiences, regardless of participants' levels of pre-task prejudice or type of argument they were exposed to, seemed to relate to their resistance or acceptance of influence. In particular, participants who had negative or positive experiences with the target group, prior to their participation, tended to oppose counter-experiential views expressed by other group members during the discussion. This supports the ELM, suggesting that personal experiences in counter-attitudinal conditions result in central processing and therefore resistance to influence (Petty & Cacioppo, 1986b).

3.5.7 Task Engagement

Finally, in terms of task engagement, this appeared to be related to participants' pre-task levels of prejudice towards the target groups. In particular, despite holding the

least socially acceptable attitudes, participants with high levels of prejudice towards the target groups indicated the highest levels of perceived task engagement compared to both participants with low or intermediate levels of pre-task prejudice. One explanation for this finding might be that those with high levels of prejudice were able to express their socially unacceptable views freely and without feeling the risk of being judged or socially rejected (DeKoster & Houtman, 2008). As a result, the ability to disclose and compare similar views, which are usually deemed socially unacceptable (DeKoster & Houtman, 2008) could thus have increased participants' perceived level of engagement in the task.

3.5.8 Limitations

3.5.8.1 Ecological validity and realism. As with experimental research in general, one of the limitations of this study relates to its lack of ecological validity. In particular, although the design attempted to mimic a naturalistic set-up, participants still knew that they were participating in a psychology study and might otherwise never have chosen to participate in these specific online discussions. Additionally, this study failed to obtain participants' demographic details. The decision to omit personal information from the data collection was made given the sensitive nature of this study and thus to increase participants' sense of anonymity.

Moreover, participants were not specifically asked about their perceived realism of the discussion or whether they guessed that there was a confederate. Yet, a number of participants emailed the researcher immediately after their participation to express their enjoyment of the task. On several occasions, this also resulted in them indicating that they did not guess that there was a confederate during the discussion until they read the debrief information. Consequently, these responses are encouraging indicators that (at least) a number of participants perceived the discussion as realistic.

3.5.8.2 Time restrictions. The findings obtained in this study are limited to the immediate time after the discussion and thus do not allow for inferences about any long-term effects. In particular, although participants with intermediate levels of prejudice indicated higher levels of prejudice immediately after taking part in the online discussions, it remains unclear whether this effect would persist over time. Specifically, previous research investigating the endurance of online influence through narrative and explicit messages on hate web pages has already demonstrated that persuasive effects can either increase or decay over time (e.g. Lee & Leets, 2002). Moreover, Lee and Leets (2002) found that both implicit and explicit low narrative content elicited persistent persuasive effects which even increased over time. They also found that explicit high narrative content (i.e. stories with plots and main characters) elicited persistent persuasive effects. Furthermore, Lee and Leets (2002) found that persuasive effects elicited by implicit high narrative content decayed over time. It would therefore be interesting for future research to explore whether the persuasive effects of online discussions (i.e. explicit, linked content) would persist over time.

3.5.8.3 Information processing. Despite indications of participants processing the information centrally or peripherally, it is not possible to make definitive inferences about participants' processing routes from this study, as this was not specifically measured. Previous research investigating persuasive strategies in the context of online hate (e.g. Lee & Leets, 2002) used a thought-generation task to explain findings relating to participants' processing routes regarding their acceptance or resistance to racist online content (Petty & Cacioppo, 1986a). A thought-generation task measures cognitive responses (i.e. central processing), in the form of expressions of favourable or unfavourable thoughts, following exposure to pro- or and counter-argumentative message content. In this study, participants also displayed some pro- and counter-arguments during the discussion task, thus

suggesting that some central processing took place. This is supported by the ELM which suggests that counter-attitudinal arguments are likely to result in participants paying close attention to the information they are presented with (i.e. central processing), given that it runs contrary to their existing views (Petty & Cacioppo, 1986b). Yet, in order to make more definitive claims about participants' processing strategies, future research should implement such measures as a thought-generation task.

3.5.8.4 Target group's 'specificness'. Finally, this study explored changes in peoples' prejudice towards a specific ethnic group, following 'like-minded' online discussions. As a result, the findings cannot explain online influence in general.

3.5.9 Future Directions

3.5.9.1 Inoculation treatments. Given that participants with intermediate levels of prejudice were the only ones influenced by online interactivity, this group presents an interesting avenue for future research investigating the influence and/or prevention of online hate. In particular, future research should investigate the effectiveness of inoculation treatments delivered via institutions, such as schools and universities, to counteract online influence. Inoculation treatments involve exposing people to pre-emptive messages about a target in order to strengthen existing attitudes/views and thus prevent persuasion (McGuire, 1961a; 1961b). Inoculation treatments have already been found to prevent persuasive attempts in a range of contexts, including politics (e.g. Pfau & Burgoon, 1988; Pfau, Kenski, Nitz & Sorenson, 1990), health campaigns (e.g. Godbold & Pfau, 2000; Parker, Ivanov & Compton, 2011; Pfau & Van Bockern, 1994), marketing (e.g. Compton & Pfau, 2004), public relations (Pfau, Haigh, Sims & Wigley, 2007; Wan & Pfau, 2004) and social networking (Compton & Pfau, 2009).

3.5.9.1.1 Education. Inoculation treatments, coupled with education about some of the consequences of online hate, delivered to students in schools and universities might

therefore present a valuable method in preventing the spread of online hate. In particular, strategies to reduce prejudice against alternative subcultures and other ethnic minorities (e.g. Roma and Travellers), by educating children, teenagers, and young adults about the possible consequences of prejudice, are already being applied in schools and universities (e.g. The Sophie Lancaster Foundation, Holocaust Educational Trust).

3.5.9.1.2 Attitude strength. In line with the findings of this study, inoculation treatments might become especially relevant for online interactions, including persuasive attempts exerted by hate groups with an audience holding weak attitudes towards the issue at hand. In particular, this study not only observed that online influence was strongest for those with weak attitudes (i.e. intermediate levels of prejudice towards the target group), but also that most recruited participants (i.e. 50%) indicated intermediate levels of prejudice (i.e. weak attitudes) towards the target group. This finding not only supports James' (2014) argument that many people hold prejudices towards Roma and Travellers, it also suggests that many people could potentially be at risk of online influence, if this number also reflects the number of people with weak attitudes towards other target groups.

3.5.9.2 Online anonymity. Future research could also investigate whether online influence in the present set-up is affected by reduced anonymity (e.g. interacting in groups over web-cams with and without sound/speech). In particular, whilst the present study did not manipulate anonymity per se (i.e. all participants were anonymous to each other), it would be interesting to see if online influence is increased/decreased through reduced anonymity. Existing research has suggested that online anonymity can lead to increased negative online behaviour (e.g. Coffey & Woolworth, 2004; Hardaker, 2010; Suler, 2004) compared to being identifiable (i.e. face-to-face interactions; Coffey & Woolworth, 2004). It is therefore likely that reduced online anonymity would lead to reduced online influence.

3.5.9.2.1 Pre-task relationships. Finally, and as pointed above, it seems unsurprising that participants who indicated little influence by other group members and felt little commitment to the group resisted online influence (Cialdini & Trost, 1998; Deutsch & Gerard, 1955; Hewstone & Martin, 2012; Kelman, 1958). Participants' low levels of group identification and commitment could be attributed to two particular factors of the design (i.e. participants' anonymity to each other and the limited time to get to know each other). Consequently, and although participants in the discussion groups held similar levels of prejudice and thus were categorised as 'like-minded', they might not have perceived other participants as 'like-minded' group members.

Future research could thus not only explore the effects of reduced online anonymity on online influence, but also consider people's pre-task relationship (i.e. knowing/liking each other). In particular, previous research has already demonstrated that online interactions between hate group members (i.e. people who belong and identify as members of these groups) can reinforce negative stereotypes about disliked out-groups (e.g. Angie et al., 2010; DeKoster & Houtman, 2009; Glaser, et al., 2002; Green, et al. 1999; Price et al., 2006; Sunstein, 2001, 2009; Wojcieszak, 2010). It is therefore most likely that knowing and liking each other (i.e. stronger group identification and commitment; Reicher et al., 2005; Turner, 1985, 1991; Turner et al., 1987) would result in increased online influence.

As such, the general rejection that social influence does not facilitate online influence (i.e. Hypothesis 2) seems too simplistic. Instead, it seems more appropriate to argue that social influence may facilitate online influence if group identification and group commitment are strong, instead of dismissing the influential effects of social influence altogether. In addition, reduced online anonymity amongst people who like each other could also lead to increased online influence.

3.5.10 Conclusion

This study builds on the small body of existing research investigating the persuasive power of online interactions, and in particular changes in prejudice leading to the development of online hate. Whilst the results of this study revealed that the majority of participants resisted online influence and thus did not change their levels of prejudice after interacting with ‘like-minded’ others, there was one group of participants who were influenced by it: participants with weak attitudes (i.e. intermediate levels of prejudice) increased their levels of prejudice towards a specific target group after they participated in an online discussion in which negative prejudice about these groups was induced.

3.6 References

- Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorisation and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology, 29*, 97-119. DOI: 10.1111/j.2044-8309.1990.tb00892.x
- Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery*. New York, NY: ACM.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Erlbaum.
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology, 41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Anti-Defamation League (2005). *Attitudes toward Jews in twelve European countries* (May 2005). Retrieved from www.adl.org:
http://www.adl.org/israel/Eur_Poll_Israel_May_2005.pdf
- Association of Chief Police Officers (2014). *ACPO hate crime manual*. ACPO: London.
- BBC News (2016, January 27). Quarter of people have witnessed hate crime, poll suggests. *BBC News*. Retrieved from <http://www.bbc.com/news/uk-35415988>

- Beirich, H. (2014). White homicide worldwide: Stormfront, the leading White supremacist web forum, has another distinction — murder capital of the internet. *Southern Poverty Law Centre Intelligence Report*. Retrieved from <http://www.splcenter.org/sites/default/files/downloads/publication/white-homicide-worldwide.pdf>
- Blazak, R. (2001). White boys to terrorist men: Target recruitment of Nazi skinheads. *American Behavioral Scientist*, 44, 982–1000. DOI: 10.1177/00027640121956629
- Bostdorff, D. M. (2004). The internet rhetoric of the Ku Klux Klan: A case study in website community building run amok. *Communication Studies*, 55, 340-361. DOI: 10.1080/10510970409388623
- British Psychological Society (2013). *Ethics guidelines for internet-mediated research*. Retrieved from <http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf>
- Burgoon, M. (1990). Language and social influence. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 51-72). New York: John Wiley.
- Charteris-Black, J. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17, 563-581.
DOI: 10.1177/09579265060666345
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 151-192). Boston, MA: McGraw-Hill.
- Coffey, B., & Woolworth, S. (2004). “Destroy the scum, and then neuter their families:” The web forum as a vehicle for community discourse? *The Social Science Journal*, 14(1), 1-14. DOI: 10.1016/j.soscij.2003.10.001

- Compton, J., & Pfau, M. (2004). Use of inoculation to foster resistance to credit card marketing targeting college students. *Journal of Applied Communication Research*, 32, 343-364. DOI: 10.1080/0090988042000276014
- Compton, J., & Pfau, M. (2009). Spreading inoculation: Inoculation, resistance to influence, and word-of-mouth communication, *Communication Theory*, 19, 9-28. DOI: 10.1111/j.1468-2885.2008.01330.x
- Cooper, J., Kelly, K. A., & Weaver, K. (2003). Attitudes, norms and social groups. In M. A. Hogg & S. Tindale (Eds.), *Blackwell handbook of social psychology: group processes*. Oxford: Blackwell.
- DeKoster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me'. *Information, Communication & Society*, 11, 1155-1176. DOI: 10.1080/13691180802266665
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of Abnormal and Social Psychology*, 51, 629-636. DOI: 10.1037/h0046408
- Douglas, K. (2007). Psychology, discrimination and hate groups online. In A. Joinson, K. McKenna, T. Postmes & U. D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 155–163). New York: Oxford University Press.
- Douglas, K. M., McGarty, C., Bliuc, A. M., & Lala, G. (2005). Understanding cyberhate: Social competition and group creativity in online White supremacy groups. *Social Science Computer Review*, 23, 68-76. DOI: 10.1177/0894439304271538
- Eagly, A. H. (1974). Comprehensibility of persuasive arguments as a determinant of opinion change. *Journal of Personality and Social Psychology*, 29, 517-554.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

- Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology, 29*, 371-389. DOI: 10.1002/(SICI)1099-0992
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behaviour: The MODE model as an integrative framework. *Advances in experimental social psychology, 23*, 75-109. DOI: 10.1016/S0065-2601(08)60318-4
- Foxman, A. H., & Wolf, C. (2013). *Viral hate containing its spread on the internet*. New York: Palgrave Macmillan.
- Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy, 3*, 29-44. DOI: 10.1111/j.1530-2415.2003.00013.x
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues, 58*, 177-193. DOI: 10.1111/1540-4560.00255
- Godbold, L. C., & Pfau, M. (2000). Conferring resistance to peer pressure among adolescents: Using inoculation theory to discourage alcohol use. *Communication Research, 27*, 411-437. DOI: 10.1177/009365000027004001
- Graham, L. (2016, February 9). 24% of young people targeted by online abuse in the last year. *CNBC*. Retrieved from <http://www.cnbc.com/2016/02/09/24-of-young-people-targeted-by-online-abuse-in-the-last-year.html>
- Green, D. P., Abelson, R. P., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice & D. T. Miller (Eds.), *Cultural divides*, (pp. 429-464). New York: Russell Sage Foundation.

- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture*, 6, 215-242. DOI: 10.1515/jplr.2010.011
- Hewstone, M., & Martin, R. (2012). Social influence. In M. Hewstone, W. Stroebe & K. Jonas (Eds.), *An introduction to social psychology* (5th ed.), (pp. 235-272). Chichester: John Wiley & Sons Ltd.
- Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorisation: A motivational theory of social identity processes. *European Review of Social Psychology*, 11, 223-255. DOI: 10.1080/147927720043000040
- Home Office (2009). *The cross-government action plan*. London: Home Office.
- James, Z. (2014). Hate crimes against Gypsies, Travellers and Roma in Europe. In N. Hall, A. Corb, P. Giannasi & J. Grieve (Eds.), *The Routledge international handbook on hate crime* (pp. 237-248). London: Routledge.
- Karasawa, M. (1991). Toward an assessment of social identity: The structure of group identification and its effects on ingroup evaluations. *British Journal of Social Psychology*, 30, 293-307. DOI: 10.1111/j.2044-8309.1991.tb00947.x
- Kelman, H. C. (1958). Compliance, identification and internalization: Three processes of attitude change, *Journal of Conflict Resolution*, 2, 51-60.
- Klein, O., Spears, R., & Reicher, S. (2007). Social identity performance: Extending the strategic side of SIDE. *Personality and Social Psychology Review*, 11, 28-45. DOI: 10.1177/108868306294588
- Krauss, R. M., & Chiu, C. (1998). Language and social behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey, *Handbook of social psychology* (4th ed., Vol. 2, pp. 41-88). New York: Oxford University Press.

- Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology*, *65*, 1132-1151. DOI: 10.1037/0022-3514.65.6.1132
- Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist*, *45*, 927-957. DOI: 10.1177/0002764202045006003
- McGuire, W. J. (1961a). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, *24*, 184-197. DOI: 10.2307/2786067
- McGuire, W. J. (1961b). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, *63*, 326-332. DOI: 10.1037/h0048344
- McNamee, L. G., Peterson, B., L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs*, *77*, 257-280. DOI: 10.1080/03637751003758227
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, *12*, 125-135. DOI: 10.1037/h0027568
- Mutz, D. (2006). *Hearing the other side: Deliberate versus participatory democracy*. New York: Cambridge University Press.
- Mutz, D. C., & Martin, P. S. (2001). Facilitating communication across lines of political difference: The role of mass media. *American Political Science Review*, *95*, 97-114.
- Myers, D. G., & Bishop, G. D. (1970). Discussions effects on racial attitudes. *Science*, *169*, 778-779. DOI: 0.1126/science.169.3947.778

- O’Keefe, D. J. (2004). Trends and prospects in persuasion theory and research. In J. S. Seiter & R. H. Gass (Eds.), *Perspectives on persuasion, social influence and compliance guiding* (pp. 31-43). Boston, MA: Person Education, Inc.
- Organisation for Security and Co-operation in Europe (2010). *Incitement vs. Freedom of Expression: Challenges of combating hate crimes motivated by hate on the Internet*. Report of the OSCE-ODIHR Expert Meeting in Warsaw. Retrieved from <http://www.osce.org/odihr/68750>
- Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.
- Parker, K. A., Ivanov, B., & Compton, J. (2011). Inoculation's efficacy with young adults' risky behaviors: Can inoculation confer cross-protection over related but untreated issues? *Health Communication, 27*, 223-233. DOI: 10.1080/10410236.2011.575541.
- Parrott, R. L. (1995). Motivation to attend to health messages: Presentation of content and linguistic considerations. In E. Maibach & R. L. Parrott (Eds.), *Designing health messages* (pp. 7–23). Thousand Oaks, CA: Sage.
- Perry, B. (2000). “Button-down terror”: The metamorphosis of the hate movement. *Sociological Focus, 33*, 113–131. DOI: 10.1080/00380237.2000.10571161
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: William C. Brown.
- Petty, R. E., & Cacioppo, J. T. (1986a). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.
- Petty, R. E., & Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, (pp. 123-205). New York: Academic Press.

- Petty, R. E., & Cacioppo, J. T. (1990). Involvement and persuasion: Tradition versus integration. *Psychological Bulletin*, *107*, 367-374. DOI: 10.1037/0033-2909.107.3.367
- Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*, 847-855. DOI: 10.1037/0022-3514.41.5.847
- Petty, R. E., Haugtvedt, C. P., & Smith, S. M. (2014). Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant and predictive of behaviour. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 93-130). Hove, East Sussex: Psychology Press.
- Petty, R. E., & Krosnick, J. A. (2014). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1-24). Hove, East Sussex: Psychology Press.
- Pfau, M., & Burgoon, M. (1988). Inoculation in political campaign communication. *Human Communication Research*, *15*, 91-111. DOI: 10.1111/j.1468-2958.1988.tb00172.x
- Pfau, M., Haigh, M. M., Sims, J., & Wigley, S. (2007). The influence of corporate front-group stealth campaigns. *Communication Research*, *34*, 73-99. DOI: 10.1177/0093650206296083
- Pfau, M., Kenski, H. C., Nitz, M., & Sorenson, J. (1990). Efficacy of inoculation strategies in promoting resistance to political attack messages: Application to direct mail. *Communication Monographs*, *57*, 25-43. DOI: 10.1080/03637759009376183
- Pfau, M., & Van Bockern, S. (1994). The persistence of inoculation in conferring resistance to smoking initiation among adolescents: The second year. *Human Communication Research*, *20*, 413-430. DOI: 10.1111/j.1468-2958.1994.tb00329.x

- Price, V., Nir, L., & Capella, J. N. (2006). Normative and informational influence in online political discussions. *Communication Theory, 16*, 47-75. DOI: 10.1111/j.1468-2885.2006.00005.x
- Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior, 29*, 610-615. DOI: 10.1016/j.chb.2012.11.011
- Reicher, S., Haslam, S. A., & Hopkins, N. (2005). Social identity and the dynamics of leadership: Leaders and followers as collaborative agents in the transformation of social reality. *The Leadership Quarterly, 16*, 547-568. DOI: 10.1016/j.leaqua.2005.06.007
- Rohlfing, S., & Sonnenberg, S. (2016). "Who is really British anyway?": A thematic analysis of responses to online hate materials. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10*. DOI: 10.5817/CP2016-4-2
- Spears, R., & Tausch, N. (2012). Prejudice and intergroup relations. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An introduction to social psychology* (5th ed., pp. 449-498). Chichester: Wiley.
- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano Jr., L. A., Benton, E. & Steinfeldt M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology, 16*, 362-371. DOI: 10.1037/a0018692
- Stern, K. S. (2002). *Hate and the internet*. New York: The American Jewish Committee.
- Stroebe, W. (2012). Strategies of attitude and behaviour change. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An introduction to social psychology* (5th ed., pp2-3-231). Chichester: Wiley.

- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7, 321-326.
DOI: 10.1089/1094931041291295.
- Sullivan, A. (1999, September 26). What's so bad about hate. *New York Times*. Retrieved from <http://www.nytimes.com/1999/09/26/magazine/what-s-so-bad-about-hate.html?pagewanted=all>
- Sunstein, C. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.
- Sunstein, C. (2009). *Going to extremes: How like minds unite and divide*. New York, NY: Oxford University Press.
- Turner, J. C. (1985). Social categorisation and the self-concept: A social cognitive theory of group behaviour. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (2nd ed., pp. 77-122). Greenwich, CT: JAI Press.
- Turner, J. C. (1991). *Social influence*. Buckingham, UK: Open University Press.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A Self-Categorization Theory*. Oxford & New York: Blackwell.
- Wan, H. H., & Pfau, M. (2004). The relative effectiveness of inoculation, bolstering, and combined approaches in crisis communication. *Journal of Public Relations Research*, 16, 301-328. DOI: 10.1207/s1532754xjpr1603_4
- Wang, Y., Cong, G., Song, G., & Xie, K. (2010). Community-based greedy algorithm for mining top-K influential nodes in mobile social networks. *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, (pp. 1039-1048). Retrieved from <http://dl.acm.org/citation.cfm?id=1835935>
- Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media Society* 12, 637-655. DOI: 10.1177/1461444809342775

Woodall, W. G., & Burgoon, J. K. (1981). The effects of nonverbal synchrony on message comprehension and persuasiveness. *Journal of Nonverbal Behavior*, 5, 207-223.

DOI: 10.1007/BF00987460

Chapter 4

“Who is really British anyway?” A Thematic Analysis of Responses to Online Hate Materials

4.1 Abstract

This article aims to add to the relatively small body of literature on online hatred. In particular, it focuses on the role Social Networking Sites (SNSs) and Social Media Platforms (SMPs) may play in the development of polarisation, by exploring how online users respond to explicit online hate materials. Specifically, this article discusses the ways in which a self-selected sample of YouTube users responded, via posting online comments, to a video clip in which a White female train passenger (called Emma) could be seen to racially abuse other passengers. Thematic analysis of the YouTube comments identified four main themes: (1) *Making Sense of Emma*, which encapsulated posters' attempts to find explanations for Emma's behaviour; (2) *Meeting Hatred with Hatred*, which described posters' attempts to oppose Emma's racism by means of resorting to aggressive, hateful language; (3) *Us versus Them*, which encapsulated posters' tendencies to categorise themselves and other posters into in- and outgroups, based on their particular stance on racism; (4) *Contesting Britishness*, which expressed posters' attempts to articulate (and contest) what it means to be British. Whilst this analysis provides some evidence that hateful web content can fuel aggressive and hateful responses, many of the comments analysed here emphasised common group membership, alongside people's right to claim membership in a particular social category (i.e. Britishness). The current evidence, therefore, suggests that, at least in the specific context of this study, hateful web content may not necessarily lead to an automatic endorsement or escalation of hatred.

4.2 Introduction

On the 21st of March 2016, 3,3 billion people worldwide were online (Internet Live Stats). The relatively recent emergence of Social Networking Sites (SNSs) has contributed noticeably to the growing numbers of Internet users. For example, Facebook had 1.6 billion users in March 2016 (Internet Live Stats), and YouTube claimed, that in 2015, one billion users visited its page per month (YouTube, n.d.).

Despite their immense popularity, some SNSs have recently come under critical scrutiny. In Britain, for example, there is increasing governmental concern about the Internet's role in the incitement of hatred, including racial hatred. There have also been claims by some researchers (e.g. Foxman & Wolf, 2013; Perry, 2000; Stern, 2012), advocacy groups, Internet Service Providers (ISPs), and charities that websites promoting hatred may contribute to the development and exacerbation of radical attitudes and behaviours and, as such, foster terrorist sympathies. However, to date, claims regarding the persuasive power of the Internet remain largely based on anecdotal observations, common sense assumptions, and on government concerns regarding the protection of members of the public.

The existence of hateful materials online seems indisputable (e.g. it can be evidenced by certain key word searches). In addition, online hate materials, including propaganda video clips on websites, such as YouTube, are regularly found on computers seized from terrorist suspects (e.g. the Boston Bombers, Robi, 2013; or the Lee Rigby murderers, Swinford, 2013). Consequently, lawmakers across the globe have begun to amend and develop legislation to account for the evolution of the Internet into a potential medium for stirring up and promoting hatred. In the United Kingdom alone, this legislation includes the Offences Against the Person Act, 1861, the Public Order Act 1986, the Malicious Communications Act, 1988, the Protection from Harassment Act 1997, the

Crime and Disorder Act 1998, the Communications Act 2003, the Criminal Justice Act 2003 and the Racial and Religious Hatred Act 2006. Consequently, in the United Kingdom, several arrests and convictions have been made in relation to the publication and distribution of online materials with the intent to incite racial hatred Colin White (BBC News, 2016); Ihjaz Ali, Kabir Ahmed & Tazwan Javed (BBC News 2012); Joshua Bonehill-Paine (Jewish News Online, 2015); Matthew Doyle (Burrows, 2016); Teenager (Crown Prosecution Services, 2010). Arrests have also been made in Australia, China, Dubai, Russia and Kazakhstan where several people have been convicted of inciting hate via social media, including ethnic hatred (e.g. Ai Takagi, see: Mail Online, 2016; Alkhanashvili, see: Kazakhstan Newline, 2015; Mukhtar, see: Radio Free Europe Radio Liberty, 2015; Pu Zhiquaing, see: Phillips, 2015; Tatyana Shevtsova-Valova, see: Radio Free Europe Radio Liberty, 2015), religious hatred (e.g. Yklas Kabduakasov, see: Spencer, 2015; UAE national, see: Agarib, 2016) and extremism (e.g. Vadim Tyumentsev, see: Reuters, 2015).

Academic research exploring online hatred has begun to emerge in several social sciences (mainly sociology and criminology). So far, this research has tended to have two main analytic foci, namely on providing (1) detailed descriptions of online hate materials or (2) characterisations of polarised individuals/groups. For example, research on online hate materials has focused on outlining the specific contents of hate-promoting websites and has provided insights into the use of propaganda to polarise and recruit new members to common ideological causes (e.g. Anti-Defamation League, 2005; Blazak, 2001; Bostdorff, 2004; Douglas, 2007; Douglas, McGarty, Bliuc & Lala, 2005, Gerstenfeld, Grant & Chiang, 2003; Levin, 2002; McNamee, Peterson & Peña, 2010; Perry, 2000; Turpin-Petrosino, 2002; Zhou, Reid, Qin, Chen & Lai, 2005). Research focusing on understanding the type of individual or group who produces hateful materials or hate

speech online has tended to concentrate specifically on people who are already polarised, thus identifying not only some personality traits associated with hate speech but also some of the ways in which polarised individuals or groups share their ideology and deliberately use prejudice to create conflict between groups (e.g. Angie et al., 2011; Cammaerts, 2009; De Koster & Houtman, 2008, Doane, 2006, Erjavec & Kovačič, 2012; Ezekiel, 2002; Glaser, Dixit & Green, 2002; Green, Abelson & Garnett, 1999; Wojcieszak, 2010).

However, thus far, research tapping into the broader *processes* underlying the development of online hatred remains limited – despite the reported increase in ‘hate blogging’, especially on SNSs and Social Media Platforms (SMPs), over time (e.g. Chau & Xu, 2007; Citron & Norton, 2011) and the recent surge in arrests and convictions for SNS/SMP-related hate offences. In particular, there is limited empirical evidence investigating how Internet users might respond to hateful online web content displayed on ordinary SNSs and SMPs (i.e. YouTube, Facebook, Twitter).¹ In particular, whilst one might expect to come across hateful/polarising content on certain websites (i.e. Stormfront, English Defence League), one might not expect to see such content during their visit to their familiar SNSs/SMPs. As a result, we know little about how Internet users might respond to or be influenced by hateful online materials displayed on SNSs/SPMs.

Many SNSs/SMPs and online news web sites now allow, and even encourage, online users to express their views or responses to different web contents (including

¹ As noted in Chapter 1, SNSs and SMPs are conceptually not the same. Yet, as YouTube’s commenting feature allows users to socially interact (i.e. communicate/network) with other chosen or random users, it shares some of the attributions of SNSs and SMPs. As a result, references made in this thesis, assigning YouTube as a SNS, are to be understood based on this overlap in attributed features.

specific statements, pictures, video clips or TV programmes) via posting/commenting. These posts or comments are usually displayed publicly and, in turn, often seem to elicit further comments/posts from other users – who might agree or disagree with either the initial poster or with the web content in question. It seems therefore plausible that certain web content could lead SNS/SMP users to respond to such content in a polarised manner. These polarised responses could, in turn, lead to (virtual) conflicts amongst SNS/SMP users.

In fact, such virtual conflicts can be observed in so-called online ‘trolling’, where certain individuals (i.e. ‘trolls’) intentionally aim to stir responses amongst users. Trolling has been attributed to online anonymity (Hardaker, 2010; Herring, Job-Sluder, Scheckler & Barab, 2002), which, in turn, links to the so called online disinhibition effect (ODE; Suler, 2004). In particular, the ODE is thought to be the result of a lack of social constraints mediated, predominantly, by online anonymity (Joinson, 1998; Suler, 2004). Whilst not all behaviour associated with online disinhibition has to be negative (i.e. increased self-disclosure), as noted above, online disinhibition can encompass negative communication, such as offensive language, harsh criticisms, anger, hatred and threats (Suler, 2004). Suler (2004) referred to this form of the ODE as ‘toxic’ online disinhibition.

Accordingly, Coffey and Woolworth (2004) found that whilst people’s responses to an alleged racially motivated murder reported in an anonymous online newspaper discussion board resulted in expressions of insults, racism, abuse and hatred (i.e. polarisation), no such responses were observed when the incident was discussed in a face-to-face meeting. They consequently attributed these polarised responses expressed online to online anonymity. In addition, Peddinti, Ross and Cappos (2014) found that compared to identifiable Twitter users, anonymous users were less inhibited in relation to their Twitter activity. They also found that there are more anonymous Twitter users following sensitive

topics, such as pornography, than identifiable ones.

Further, Doughty, Lawson, Linehan, Rowland and Bennett (2014) found that controversial television broadcasts elicited many polarised (i.e. anti-social, abusive, negative, aggressive and hostile) tweets - a phenomenon frequently referred to as 'flaming' (O'Sullivan & Flanagin, 2003; Suler, 2004; Turnage, 2008). In Doughty et al.'s (2014) study 'flaming', which is described as a negative aspect of the ODE, tended to be directed at the individuals or specific communities depicted in these broadcasts. In addition, Lange (2007) and Moor, Heuvelman and Verleur (2010) investigated possible causes of 'flaming' behaviour on YouTube and found that people predominantly 'flamed' to express their disagreement with certain video or comment content. They also found that a small number of users 'flamed' for mere entertainment (Moor et al., 2010) and enjoyment (Lange, 2007). Lange (2007) and Moor et al.'s (2010) findings also indicate that, although 'flaming' seems prevalent on YouTube, it is not always perceived to be harmful, even by the recipients or 'targets' of 'flames'.

Moreover, Faulkner and Bliuc (2016) explored discursive strategies of racist and anti-racist supporters expressed on online news web sites. They found that moral disengagement is a very common strategy used by those supporting racism in order to justify or 'feel okay' about their views. They also found that those who opposed racist views/acts rarely used this strategy when expressing their views online. Given that expressions of racial attitudes are no longer socially acceptable (Crowne & Marlowe, 1960; Furnham, 1986; Gawronski et al., 2007; Joinson, 1999; Steinfeldt et al., 2010), it does not seem surprising that moral disengagement is predominantly observed in those supporting racist attitudes.

Current evidence therefore presents some strategies through which the expression of certain attitudes (i.e. racial) can be justified. In addition, it suggests that SNS/SMP

content may not only facilitate discussions amongst online users but can also lead to attitude polarisation. However, to what extent does controversial web content succeed in influencing or changing people's actual attitudes or views?

Steinfeldt et al. (2010) found that online newspaper forums often provide misinformation about specific groups (i.e. American Indians, in their case) which, in turn, can reinforce existing negative stereotypes and thus facilitate the expression of racist attitudes. More specifically, Lee and Leets (2002) argued that White supremacist web content can influence attitudes and views of polarised as well as 'neutral' adolescent online users. Furthermore, Rauch and Schanz (2013) found that, in the context of White supremacist online content, frequency of Facebook use influenced users in one of two ways. In particular, compared to infrequent Facebook users, frequent users were both more likely to agree with or reject racist messages (e.g. about White superiority), thus suggesting that those who frequently use Facebook are at higher risk of responding in a polarised manner to its content compared to those who infrequently use it.

In sum, the evidence outlined above suggests that the content of materials posted on online forums or SNSs/SMPs can influence and polarise the expression of people's views as well as lead to online abuse— even in cases where the initial web content is not specifically problematic or hateful. Yet, the extent to which explicit online hate materials might elicit hateful responses in online users more generally remains unexplored. For example, given recent media coverage (e.g. Castle, 2013; McNab, 2015; Nirvana News, 2016; Poch, 2015; Risen, 2014) focusing on the use of YouTube and other social networking sites (e.g. for promoting domestic violence or for terrorist propaganda and recruitment), it seems increasingly likely that people come across hateful content without necessarily seeking to do so. This then, raises the question of how SNS/SMP users in general might respond to *explicit* hate materials.

This study therefore aims to explore the following research question: *How do YouTube users respond to online materials that can be described as explicitly inciting hatred?*

The research question here is deliberately broad in order to allow for a qualitative, bottom-up and inductively-driven analytic approach – which has been deemed the most appropriate research strategy for under-researched areas (Braun & Clarke, 2006). It is therefore important to stress that this study aims to offer an in-depth understanding of people's responses to a very specific type of online hate material (i.e. a video clip with racist content) within a particular social networking context (i.e. YouTube). In other words, this study seeks to provide a detailed description of the qualitative nature of people's online responses in a specific context – and does not set out to make any claims about online responses to hateful materials in general. However, the author hopes that the qualitative findings outlined below will provide a springboard for further quantitative research in this field to evaluate the extent to which the key findings can be generalised.

4.3 Method

4.3.1 Materials

4.3.1.1 Data source. This analysis was based on comments made by users of the Social Networking Site (SNS) *YouTube* in response to a particular video clip uploaded on its website. The video clip depicted a White woman who racially abused other passengers. The woman, later identified as Emma West, was travelling with her toddler son on the London underground in October 2011. The scene, which lasted approximately five minutes, was filmed by a passenger on his/her mobile phone, who subsequently uploaded it on 28th November 2011 as video clip onto YouTube (for a detailed description of the clip's content, please see below). The clip prompted an unprecedented number of viewers to respond to its content via YouTube's commenting feature – and these responses

provided the raw data for this study. It should be noted that some of these comments contain offensive, racist and potentially upsetting content.

The nature of our choice of analytic materials was therefore opportunistic. In particular, when the video clip in question went viral online, the author became aware of the analytic potential the accompanying comments presented for addressing the current research question. To the author's knowledge, the video clip was the first of its kind (in the UK) to have elicited such large numbers of views and comments. Using this data source, therefore, had several methodological advantages. First, the voluntary exposure of YouTubers to the clip's content circumvented any ethical concerns an experimental design (i.e. the experimental use of hateful stimuli) would have posed. Second, users' voluntary choice to comment on the clip's content was likely to render responses more ecological valid than responses elicited in experimental settings. Third, unlike in many laboratory experiments, there was no researcher engagement with 'respondents', thus circumventing experimenter bias or social desirability effects.

The time-frame for data collection, in turn, was determined by the sheer volume of online comments made in response to the YouTube clip. Forty-eight hours after the clip's upload, the author decided that the volume of comments was more than sufficient whilst still presenting a manageable analytic task. Also, the original video clip was removed at approximately the same time because the police started a criminal investigation of the woman depicted in the clip.

Consequently, the original raw data consisted of 71,295 comments. The actual content of the video clip is described below, followed by an outline of the different types of YouTubers' responses the clip elicited after its upload, as well as the legal consequences Emma West (i.e. the woman depicted in the video clip) faced for her behaviour.

4.3.1.2 Ethical considerations. This study adheres to the British Psychology

Society's (BPS) Ethics Guidelines on Internet-Mediated Research (IMR, 2013).

Specifically, the online comments that constituted the data source for this study were available within the public domain on a large SNS/SMP. As such, the responses were visible and freely accessible to anyone using YouTube. In addition, YouTube's terms of service explicitly state that, by uploading/posting content (including comments) onto the site, users automatically consent to this content being used (i.e. reproduced, distributed, prepared in derivative works) by third parties. Whilst YouTubers' right to withdraw their responses was restricted in this study (i.e. users did not specifically know that their comments would be used in this study) all user names were anonymised, ensuring their anonymity. Moreover, the benefit of the observational (i.e. naturalistic) and unobtrusive approach avoided potentially feelings of unwelcome intrusions amongst YouTubers.

4.3.1.3 The YouTube video clip. At the start of the video clip Emma West can be heard posing the rhetorical question "*...what has the UK come to...*" - which she then proceeds to answer herself by stating "*...a load of black people and a load of fucking Polish*". Emma then verbally attacks several fellow passengers for "*not being English*", as well as telling them to "*go back to their own country*". Next, a Black woman asks Emma to refrain from swearing for the sake of the children present on the train. In response, Emma demands the Black woman state which country she is from, to which the Black woman responds that she, too, is English. Emma appears not to believe her. The clip continues by depicting Emma telling the same woman to "*go back*" to her "*own country*". Emma then attempts to specify which country the Black woman should "*go back*" to; she is, however, unable to articulate this correctly. In particular, she refers to an awkward combination of the two countries Nicaragua and Siberia, which in this specific combination sounds to other passengers like the racially offensive term NIGGER. This, in turn, can be seen to upset a young Black man behind Emma so much that he stands up, visibly agitated, in

response to which two fellow passengers immediately get up and quickly calm him down; he then returns to his seat. Shortly after, a young White woman can be seen to rush towards the upset young Black man and embrace him. During this, Emma can be seen to argue with other female passengers calling out “*You ain’t fucking British! You are Black!*” Then a baby’s cries can be heard, to which a seemingly angry White woman steps forward and shouts at Emma to “*shut up*” for waking up her baby. The woman can then be heard stating that she is English and asking Emma what she has to say to *her*. Whilst the argument between the two women continues, the video clip finishes.

4.3.1.4 Online YouTube responses. The video clip went viral after its upload and was viewed on YouTube more than 3 million times within the first 48 hours of its upload. Viewers’ responses included comments making death threats towards Emma, her child, and other YouTubers, general threats of violence, as well as offensive and racist comments, which resulted in YouTube’s removal of the clip from their website after approximately 48 hours. Since then, Emma West was arrested and charged with a racially aggravated public order offence. Emma pleaded guilty to racially aggravated harassment, alarm or distress. Additionally, several copies of the video clip have re-emerged on YouTube, resulting in an estimated 11 million views to date, including repeat viewings.

4.3.2 Analytic Strategy

4.3.2.1 Data preparation. YouTube comments to the video clip were first imported into Microsoft Word and then formatted uniformly (Arial, size 12 with single line and one paragraph spacing between posts) to provide visual ease during the analytic process. This also aided initial familiarisation with the content. Next, comments that solely consisted of Computer-Mediated-Communication (CMC) abbreviation acronyms (Baron, 2004) such as *LOL*, and response tokens such as *yeah*, were deleted if the connection or relevance to other comments could not be easily established. Additionally, repeat

comments made by the same users (i.e. displaying the same verbatim content) and comments made in a foreign language (30 in total) were deleted, leaving approximately 71.000 comments for analysis.

4.3.2.2 Preliminary analysis. Initially, the comments seemed to be too short and too many in number to allow for any meaningful in-depth qualitative analysis. The data were therefore initially approached from a content analytic perspective. Based on the nature of the video clip's content, the research interest in responses to online hatred, and some previous findings (Coffey & Woolworth, 2004; Doughty et al., 2014), the researcher expected to find different levels of escalation (i.e. aggression, abuse, racism). In particular, the data was approached by manually colour-coding any responses relating to levels of escalation in an attempt to aid visual identification and management within the large data set.

However, the results of this initial content-based coding revealed no consistent evidence of levels of escalation as such. Instead, and despite the brevity of individual responses, the comments seemed richer than first anticipated, which was evidenced by the emergence of more and more content based codes. In particular, the expanded colour-coding process resulted in a total of 28 codes, including general expression of aggression, hostility, threats, violence, blame and justification for the *rant*, as well as generally tapping into different stereotypes, and general descriptions of Emma West (see Appendix XIII).

Based on the initial coding scheme, the researcher manually coded the entire data corpus, whilst three research assistants each manually coded approximately one third of it. The coding was then compared and discussed, which revealed high inter-rater reliability.

Nevertheless, the large number of emergent content codes, suggestive of an unanticipated richness of the data, led to the decision to expand the analytic approach from a content analytic to a thematic one. Therefore, the analysis presented below is based on the

latter thematic analytic approach.

4.3.2.3 Data analysis. Based on the relative lack of existing research in this area, as well as the absence of an overall theoretical framework with which to account for how online users respond to online hate materials on SNSs/SMPs, a bottom up, inductive approach to the data was taken. Overall, the thematic analytic process followed Allen, Bromely, Kuyken and Sonnenberg's (2009) procedure. Generally, Thematic Analysis (TA) permits for a degree of epistemological and theoretical flexibility and is therefore not bound or limited to any specific pre-existing theoretical framework - yet, it does allow for the consideration of relevant theories during the analytic process (Braun & Clarke, 2006). This flexibility thus allows for the identification of themes at both the semantic (i.e. 'micro') and the latent (i.e. 'macro') level of analysis (Allen et al., 2009, Braun & Clarke, 2006). Given the broad inductive scope of the research question, both levels of analysis are drawn on here.

The thematic analytic process involved reading and re-reading the entire data set several times to become more familiar with the data and to identify themes related to the overall research question. A theme was deemed as such if, first, it tapped into relevant aspects concerning the overall research question and, second, if it represented a response-trend across the entire data set (Braun & Clarke, 2006). Reading and re-reading the data set continued throughout the formulation of themes, ensuring that the final themes remained data driven and that the chosen YouTube extracts (i.e. comments) accurately illustrated the corresponding theme/sub-theme. Themes were compared and were either separated further to show different nuances in meaning, or grouped together according to their overlap with one another (Allen et al., 2009). This process allowed for the development of an analytic hierarchy, by ordering the data into overarching themes and corresponding sub-themes (Allen et al., 2009). An overarching theme was defined as such if it appeared to be more

inclusive (i.e. higher in the thematic hierarchy) and possessed greater analytic scope and explanatory strength than a sub-theme (Gleibs, Sonnenberg & Haslam, 2014). Overall, the aim of the thematic coding process was to identify and organise emerging themes that were internally homogenous, externally heterogeneous and had explanatory power (Allen et al., 2009). Attention was also paid to prevalence which was determined by the frequency of responses that illustrated each theme (Braun & Clarke, 2006).

4.4 Results

Overall, the majority of responses condemned the content of the video clip. Analysis of YouTube users' responses to the video identified four overarching themes: (1) *Making Sense of Emma*, (2) *Meeting Hatred with Hatred*, (3) *Us versus Them*, and (4) *Contesting Britishness* (see Table 4.1 for an overview). These main themes are examined in detail below under separate section headings. Each section includes a broad definition and a narrative description of the respective theme content as well as evidence from the actual online comments. Corresponding sub-themes will be presented as sub-headings. It should be noted that here, as in other qualitative analyses (see Gleibs et al., 2014), is a degree of overlap between themes. For example, all themes reflect how posters, to some extent, attempted to position themselves (and other posters) in relation to racism. It should also be noted that many of the extracts below contain grammatical and spelling errors – which were corrected by putting the corrections in square brackets only in some cases to help clarify the meaning of these extracts.

Table 4.1

Distribution of overarching theme and sub-theme frequencies including examples obtained from the thematic analysis.

Main Theme	Sub-Theme	Example Extract(s)	Frequency*
Making Sense of Emma**		<i>“She’s on drugs. Nobody can rationally act like that. (...)”</i> ; <i>She must be high or drunk?</i> <i>(...)”; “cocaine is a hell of a drug”</i> ; <i>“gurning her face off....cunt”</i> .	14%
Meeting Hatred with Hatred**		<i>“Im British through and through and all I can say is, this rat-faced ugly chav cunt needs to be dragged off by her skanky greasy hair and fucking shot.”</i>	10%
Us Versus Them**		<i>“Stupid Bitch! should know if it wasn't for us Non- brits, then maybe Britain wouldnt be as successful as they are. They STEAL our resources, they STEAL our oils, they STEAL our country's. And YET this British Bitch is complaining, it shouldnt be her complaining it should be the non brits complaining!”</i>	23%
	Sense of injustice**	<i>“Why is it that blacks can get away with everything like murder, rape and taking over the country... and when british try we all get banged up....i think we should all go to africa and make their country ours and make them all slaves again.... BACK TO THE GOOD OLD DAYS OF SLAVE TRADE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!”</i>	2%

Sense of deprivation	<i>“(…) white population lose out on housing jobs and benefits because of foreigners... how is that fair???”</i>	9%
Contesting Britishness**	<i>“how can people not understand that you can be British and black if you are born in Britain then you are British you can get a British passport i want to literally stab her she is a disgrace to not only Britain but human kind.”</i>	12%
Britishness under threat	<i>“This chick is badass! And she's right, 80% of that tram is not british.. Britain is dying, and she is has at least the courage to stand up for her nation.”</i>	3%
We are all immigrants	<i>“Throughout history, Britain was invaded by Germans, Danes, the French, the Italians, with migrations of gypsies from Eastern Europe. Her ancestors are probably ethnic.”</i>	4%

*The frequency of the main themes is calculated from the combined amount of comments of the main- and sub-themes.

**These extracts contain some aspects which are also encapsulated in the other themes (i.e. meeting hatred with hatred and the notion of Britishness).

4.4.1 Making Sense of Emma

Posters were obviously stirred by the content of the video clip and their responses often took a stance on racism. The video content seemed to evoke a range of negative emotions (i.e. expressions of distress) in several respondents. These emotions included shame (“*Such a rude, ignorant women makes me ashamed to be British*”), disgust and embarrassment (“*You disgust me. You are the reason I am sometimes embarrassed to admit that I'm British.*”), upset (“*This video sickens me. She needs a reality check. I'm deeply upset by this footage.*”), and shock (“*I'm actually shocked! I can't believe what I have just watched?!*”; “*I am so shocked and appalled by this woman...*”).

Notably, respondents who expressed negative emotions condemned what they saw as Emma’s overt, racist behaviour. Many posters seemed to deal with what they had seen in the video clip by attempting to make sense of Emma and her behaviour. This *sense-making* focused on Emma’s personal characteristics, and particularly on attempting to attribute Emma's behaviour to internal, intra-psychic processes (i.e. alcohol and/or drug consumption):

OMG that woman is vile. And obviously either drunk or on drugs (judging from facial expressions and slurring). Luckily we don't all feel like her and she is one of a minority of people.. (Poster A)

This extract demonstrates how many respondents who disagreed with Emma’s racist behaviour, tended to attribute her behaviour to negative personal characteristics, which, in turn, enabled them to distance themselves from Emma. This distancing is evidenced by “*we don’t all feel like her and she is one of a minority*” which, not only rejects racism but also positions the poster in opposition to Emma and other perceived racists and thus removes any basis for socially identifying with her. As will become clear later, the process of distancing also relates to another theme, namely *Us versus Them*.

The link between disapproval of Emma's behaviour and attempts to make sense of it by focusing on Emma's perceived negative personal and socio-economic characteristics becomes even clearer in the extract below. Specifically, here the poster attempts to make sense of Emma's behaviour by explicitly referring to her perceived low intelligence, low education, and low socio-economic status. Moreover, here these personal characteristics (i.e. negative stereotypes) are seen to be associated directly with racist attitudes:

The fact she can't string her opinions without saying FUCK says a lot more about her intelligence. would she happen to be on benefits and live in a council home by any chance? Its usually those people that moan about others but are happy to scrounge benefits and freebies of others who work in Britain and pay taxes like the people she insults (Poster B)

Here, there is an allusion to Emma's racism as a direct consequence of her perceived low intelligence and low socio-economic status. In particular, these attributions seem to serve primarily to label Emma as a racist – in other words, as one of "*those people*" (i.e. an out-group member). Here, this strategy to position Emma as a racist (i.e. an out-group member) further seems to serve as a means of making sense of Emma's behaviour.

In addition to explaining Emma's behaviour on the basis of a perceived lack of education, the following extract invokes a comparison between Emma and immigrants (i.e. immigrants "*have more skills*" and better work ethics than Emma). This comparison further serves as a way of discrediting Emma. The poster also tries to makes sense of Emma's behaviour by linking it to a particular lifestyle (i.e. sexual transgression and work avoidance), which has been linked with poverty (Parisi, 1998):

This is what happens when u don't get an education that informs you of the realities of immigration. immigrants of all races have to work twice as hard to get jobs here and often have many more skills, in speaking several languages. I wonder what this ignorant fuck's skills are? probably shagging around and talking about how 'blacks' have ruined the country when she probably has never held down a job in her life! (Poster C)

Making sense of Emma's behaviour was also often associated with reference to Emma's role as an unfit mother (i.e. "*a bad mother*" and "*a bad role model*"). In particular, responses frequently expressed concern for Emma's child, such as "*I feel sorry for her little boy*", "*she (i.e. Emma) used the child as a shield*", "*I hope social services will take her child away*", or "*poor child*". This, then, corresponds to broader strategies that have been observed in the media (Thomas, 1998) which aim to discredit single mothers through negative stereotyping and their portrayal as immoral and neglectful.

By referring to Emma and/or other perceived racist posters in hateful terms, comments frequently went beyond simply trying to make sense of Emma's behaviour, attempting to take a stance on racism:

Most of, if not all of the people backing this horrible cunt are unemployed dole spongers who use foreigners as an excuse for not getting a job. Waists of space racist pricks. most of the English football team are black. Show some respect (Poster D)

This, then, goes beyond trying to make sense of Emma: here, insulting and negatively stereotyping Emma and other racists serves to demonstrate the poster's opposition to their views. As will be seen, this use of hostile and aggressive language also links to the next theme (i.e. *Meeting Hatred with Hatred*). Also, the poster's reference here to the ethnic diversity of the English football team – a symbol of national pride –

foreshadows the final theme *Contesting Britishness*.

Overall, *Making Sense of Emma* encapsulated respondents' attempts to understand and explain Emma. Specifically, respondents drew on a range of negative stereotypes both on a personal level (i.e. invoking alcohol/drug abuse, low intelligence, low education, 'bad mother/role model') and a social level (i.e. invoking a low socio-economic status) in order to account for Emma's behaviour and to distance themselves from (Emma's) racist views by positioning her as 'other' (i.e. an out-group member). This process frequently involved the use of strong, hateful language as the following extract illustrates:

Its funny how this silly lady thinks its 'the blacks' that are ruining her country when she is sitting on a tram with her son clearly on drugs and carrying on and the only person decent enough to tell her not to swear in front of kids is black..... I know who I would rather live next door two, the black lady and not the coked up whore. (Poster E)

4.4.2 Meeting Hatred with Hatred

This thick bitch would have got kicked the fuk out of by me the dik head in the back that takes his bag off why didnt he do anything.. Faggot bitch would get killed lil tart if i was there i would kill the bitch !!!! FUK YU DIE BITCH (Poster F)

This extract epitomises the essence of *Meeting Hatred with Hatred* as another overarching theme in this analysis. In particular, comments encapsulated by this theme depicted posters' tendency to strongly object to Emma's behaviour by means of demeaning and dehumanising her. Specifically, many posters responded to Emma's behaviour by using hateful (including racist/abusive) and aggressive language, such as "*she has a mentality like Hitler*", "*she shouldn't be allowed to breed*", "*ugly ass throwback gobshite of a tredless bucketpussy usless eater oxygen thieving minus 80 I.Q*", "*lazy*", "*whore*",

"cunt", "chav", "slag", "anti-social, scum-sucking retard", "pikey", "albino monkey", or "honkey". This use of hateful language in response to Emma's behaviour seems paradoxical, in so far, as it reflects an attempt to counter Emma's apparent racism and hatred with further expressions of hatred (hence *Meeting Hatred with Hatred*).

Along similar lines, posters frequently dehumanised Emma by referring to her as "a piece of filth", "gutter trash", "White trash", "mosh maul", "monstrosity", "plankton", "a thing" or a "scum (bag)". These terms invoke a perceived lack of humanity on Emma's part, thus placing her outside 'normal' human society and calling for her social exclusion (e.g. into "the sewer/gutter/trash", a place where society discards its unwanted waste). The dehumanisation of Emma thus represents another way in which posters attempted to distance themselves from her and the video's content.

In addition, posters who condemned Emma's behaviour called for her punishment, albeit on a symbolical level rather than a literal one. For example, posters referred to "this lazy sponging cunt should be put down by a vet. Although animals are better looking than that scum. I suggest she goes back to where she came from the sewer". As these examples illustrate, Emma's dehumanisation and calls for her punishment frequently went hand in hand – indeed, the latter seemed predicated on the former. These calls for punishment seemed to underscore further posters' apparent opposition to racism. Further calls for Emma's punishment included "hitting her", "knocking her out", "beating her", "stabbing her", "slashing her throat", "deporting her", "locking her up in the sewer with the rats", or to "rot in hell" – which, again, highlight posters' opposition to Emma's views and behaviour.

Overall, *Meeting Hatred with Hatred* encapsulated one particular method posters used to express their disagreement with the video clip's content. In other words, in their use of aggressive, abusive, threatening, and hateful language, posters attempted to distance

themselves from Emma's expressions of racism. These attempts at distancing are also evident in the following theme.

4.4.3 Us versus Them

This prevalent theme describes how posters attempted to either distance themselves from or affiliate with Emma and her views by invoking a sense of “us” versus “them”. This, in turn, echoes previous work focussing social identity processes (e.g. Blackwood, Hopkins, Reicher, 2012, 2013, 2015; Reicher, Haslam & Rath, 2008; Tajfel & Turner, 1979; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). For example, posters condemning Emma tended to position themselves alongside other posters who shared their own seemingly anti-racist position – whilst, at the same time, distancing themselves from those who advocated or supported Emma's views (and vice versa). In particular, those posters who did not share Emma's views tended to refer to her and those who shared her views, as ‘them’ - in other words, as out-group members. Posters opposing racism also tended to refer to other like-minded YouTube users as *us*, that is, as in-group members. *Us versus Them* consisted of two sub-themes that were labelled *Sense of Injustice* and *Sense of Deprivation*.

As already indicated, many posters condemned racist views and consequently distanced themselves from Emma:

@Participant X¹: [...] All racism is wrong. This woman is an embarrassment to where I come from. Hate engenders hate. Even the biggest, most intolerant racists must recognise exceptions to their misplaced and ignorant rules. Why not try giving everyone the benefit of the doubt? I don't want to be mistaken for someone like you, or this stupid woman on the video, so why should I assume anything about anyone else? (Poster G)

This poster distances him/herself from racism in two ways: first, on a social level, by calling “*all racism wrong*” and by labelling racists, as a group, as “*intolerant*” and “*ignorant*” and second, on a personal level, by describing Emma in negative terms, calling her an “*embarrassment*”, “*stupid*”, and “*I don't want to be mistaken for* (i.e. associated with) *someone like you*” (i.e. someone with racist views). Conversely, posters supporting Emma's views attempted to distance themselves from immigrants and immigration supporters:

@Participant X²: True, that was adorable how she stood up for her race and the bravery to say what everyone was thinking in front of adversity and to all the immigrant invaders who came to enforce their foreign oppressive laws and genocide [on] the land of the indigenous Whites. (Poster H)

This comment displays categorical thinking through its use of distancing pronouns,

¹ This comment was made in response to another comment, expressing opposition to the poster's views.

² Please note that this comment was also made in response to another comment. In this instance, the comment was aimed at a poster who shared the views of this particular poster.

such as “*their*” *foreign oppressive laws*, implying that laws were not voted for or made by *us* (i.e. here, White people). Consequently, the poster here distinguishes between the in-group self (i.e. the ‘us’ supporting Emma’s views) and ‘the other’ or out-group (i.e. those opposing Emma). Specifically, the above terminology depicts ‘us’ as good, while ‘they’ are bad. Here, the contrast between negative terms, such as “*invaders*”, “*oppressive*”, “*genocide*”, “*indigenous White*”, and positive terms, such as “*adorable*”, “*stood up*” and “*bravery*” emphasises the distinction between the in-group and out-group (i.e. perceived racial differences). The term ‘genocide’ here also links to a perceived threat, ostensibly posed by immigrants, to posters’ social identity (i.e., social identity threat; Branscombe, Ellemers, Spears & Doosje, 1999; Reicher et al., 2008; Verkuyten, 2013) which will be discussed in more detail under *Contesting Britishness*.

Some comments in support of Emma not only created a sense of *Us versus Them* through use of in-/out-group defining pronouns, but also by displaying high levels of hostility towards out-group members, as illustrated in the following quote:

GOOD FOR HER! THEY CAN ALL FUCK OFF BACK TO WHERE THEY BELONG, AND THE TRAITORS WHO BROUGHT THEM TO THE U.K. MUST BE ARRESTED AND FUCKNG EXECUTED FOR TREASON!!!
FUCK OFF! FUCK OFF! FUCK OFF!!! (Poster I)

This comment clearly demonstrates the poster’s positive self-representation and negative other-representation, through out-group defining pronouns (i.e. *they/them*) and, in doing so, distances him/herself not only from ‘foreigners’ (i.e. ‘they’ do not ‘belong’ in the UK), but also from those people in the UK who support immigration (i.e. “*traitors*”, who should be “*executed for treason*”).

This hostility towards immigration supporters indicates a certain degree of conflict as the recipients of this hostility also belong to a potential *us* (i.e. the British). Here, British

nationals supporting immigration are categorised as out-group members by those opposing immigration which suggests that hostility expressed towards ostensible out-group members serves as a strategy to create inter-group conflict and justify action against them.

Overall then, *Us versus Them* tapped into broad social identity processes. That is, the theme and its two subordinate themes (see below) encapsulated posters' attempts to position themselves and other posters in line with the construction of in- and out-groups. As such, respondents' positioning frequently drew on the issue of immigration – which is exemplified by the following sub-themes, *Sense of Injustice* and *Sense of Deprivation*.

4.4.3.1 Sense of injustice. This sub-theme illustrates the way in which posters constructed a sense of *Us versus Them* by invoking immigration. In particular, those who voiced racist views tended to express their perceived sense of injustice by claiming that immigrants receive preferential treatment. Respondents who opposed racist views tended to argue that everyone “*born in Britain*” should be considered “*just as British as we (i.e. White British) are*” and therefore “*have the same rights to live in Britain*” (i.e. here, a sense of injustice consisted of denying immigrants their right to claim to be born in Britain as Britishness). These types of responses also related to notions of ‘Britishness’ and link to the final theme *Contesting Britishness*. There were also many posters who argued that immigration was an inevitable consequence of Britain's colonial history:

Hey lady - you don't like people from other countries or ethnic backgrounds in your country or calling them British or English - you guys started it! invading and looting all those other countries, staying in India, Pakistan, parts of Africa etc - now it's our turn!! PS. - rest of Britain who aren't racist like this stupid cow - we love you, we love your pubs, your beer & stouts, your football and your culture. Peace! (Poster J)

Again, here, a sense of *Us versus Them* is created through use of specific out-

group-defining pronouns (i.e. *you* and *your* to address Emma and other British people), and in-group defining pronouns (i.e. *our* and *we* in relation to members of other countries and ethnic minorities). The poster also proclaims immigration an inevitable and direct consequence of British colonialism. In particular, his/her references to *invading* and *looting* express a sense of unfairness and injustice towards people from Britain's former colonies. Whilst this also suggests that s/he identifies as non-British, the poster also distinguishes between two different versions of British identity, namely a non-racist and racist British one, thereby differentiating between different types of British national group membership.

There were, however, also posters who endorsed Emma's stance on immigration:

[...] Islamic extremists do run round screaming "kill the infidel" and the police stand by and do absolutely nothing!. One law for black and another for white isn't working, any more than a housing or education policy that discriminates against established nationals, and merely fosters disharmony. (Poster K)

Although this poster does not directly categorise him/herself as a member of a specific in-group, this comment creates a sense of *Us* versus *Them* (i.e. Islamists extremists and Black people versus White people) through ethnic stereotyping. In particular, this post implies that all Muslims are Islamic extremists and not White – a misleading fusion which is likely to reinforce existing negative stereotypes. More specifically, the comment expresses discontent concerning the British establishment (i.e. laws and government policy), whereby the British criminal justice system is accused of unjustly differentiating between groups by "*discriminating against established nationals*".

Whilst the poster in the above extract implicitly suggests that current laws/policies are unfair and contribute towards social division, other posters accused the British government and criminal justice system explicitly of committing injustices against those who are 'really British'. Specifically, their sense of injustice referred to accusations of

perceived persecution of ‘real Brits’ for ‘minor crimes’, such as "*expressing politically incorrect views*", whilst maintaining that ‘severe’ crimes, such as "*murder*", "*rape*", or "*burning poppies in the street*", committed by those ‘non-British’ (i.e. immigrants) "*remain unpunished*".

Thus, some responses to the video clip accused the British government of applying different rules to different people (i.e. favouring immigrants and discriminating against ‘the British’) - which directly related to posters’ own attempts to differentiate between groups (i.e. *Us* versus *Them*). This sense of injustice (i.e. the claim that the establishment unjustly differentiates between groups) here also relates to the next sub-theme.

4.4.3.2. Sense of deprivation. The comments falling under this sub-theme conveyed posters’ sense of feeling deprived. In particular, those holding racist and anti-immigration views tended to express a sense of deprivation via creating a sense of *Us versus Them* (i.e. here ‘us’ versus ‘foreigners’). Some posters blamed ‘foreigners’ for their own financial insecurity and unemployment, including familiar phrases, such as "*they take our jobs*", "*they are prioritised for housing*" and "*they receive special benefits and free health care*", whilst "*we lose out to them*". Consequently, and as illustrated in the following extract, posters also expressed their sense of deprivation by drawing explicit group distinctions (i.e. racial out-grouping):

I wish more people would stand up for what they believe in we have a freedom of speech but somehow this has been withdrawn with this and why??? she has an opinion and is expressing it..... if people are offended then sorry they shouldn't watch it or be here to be offended against it..... we have got too many immigrants due to legislation..... and what a load of shit it is..... white population lose out on housing jobs and benefits because of foreigners..... how is that fair??? (Poster L)

The above reference to "*the white population losing out on housing, jobs, and benefits because of foreigners...*" not only suggests a sense of deprivation in general but, particularly, a sense of 'us' being deprived compared to 'them' (i.e. see fraternal relative deprivation, Runciman, 1966). For instance, 'our' (i.e. White British) socio-economic insecurity is here, blamed on 'them' (i.e. immigrants).

Overall, the theme *Us versus Them* reflected social identity processes as posters attempted to distinguish between different in- and out-groups (e.g. Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner et al., 1987). In particular, posters attempted to align themselves with other posters whose views they shared (*Us*) and to differentiate themselves from those whose views they opposed (versus *Them*). In doing so, comments frequently touched on notions of 'Britishness' – and thus specifically on the issue of who should be included in this social category and who constitutes *Us*. This then, relates to *Contesting Britishness*, the final theme, which will be discussed next.

4.4.4 Contesting Britishness

In the majority of posts, attempts were made to define what it means to be British by constructing and contesting different versions of a British national identity. Here, posters did not simply try to position themselves in relation to Emma - instead, this theme indicates the extent to which respondents attempt to go beyond taking a specific stance on racism. In particular, this prevalent theme describes posters' ambivalence about the notion of 'Britishness' or national identity content, by capturing the ways in which comments contested the criteria by which 'Britishness' should appropriately be defined. This is perhaps unsurprising especially as the video clip showed Emma challenging the British identity of a Black fellow passenger (i.e. referring to her as non-British). Also, given the diversity of YouTube users, one might expect variations in posters' references to national identity and sense of Britishness.

Whilst endorsement of racist views and an anti-immigration stance need not necessarily go hand in hand, there frequently appeared to be a degree of overlap between the two. For example, here, respondents who positively embraced immigration did not always seem to differentiate between racism on the one hand and anti-immigration on the other. They therefore frequently referred to those who rejected immigration as racists.

Overall, respondents' accounts of the criteria defining 'Britishness' ranged from being "*multi-cultural*", "*tolerant*", "*accepting of others*", "*patriotic*", "*born in Britain*", "*White*", to "*holding a British passport*", or rejecting the existence of 'true Britishness' altogether:

She doesn't have respect. No one is PURE english no more, not even the queen and the royal family is and I bet she isn't either. I'm mixed my father is english and my mum's thai, I was born here in the UK and I am a BRITISH. I follow every thing she does. .. I do the same as she, what just because My mum's from a different country? Just because these people are different colour means there not British.. HOW SHE WAS SAYING THIS IS MY BRITAIN.. NO BITCH THIS IS OUR BRITAIN. LIVE WITH IT! (Poster M)

Here, the poster clearly rejects Emma's version of 'Britishness' and, particularly, the view that anyone can be considered 'purely' British – including the ultimate symbol of 'Britishness', namely the Royal Family (ETHNOS, 2005). In particular, s/he rejects the idea that 'Britishness' can be determined by one's skin colour. In doing so, this poster not only distances him/herself from Emma's views, s/he also implicitly affirms immigrants' and non-White people's right to claim a British identity.

The following two sub-themes relate to the degree to which the boundaries of 'Britishness' are understood as inclusive of 'others'. These were *Britishness Under Threat*, in which multiculturalism was claimed to be either the 'undoing' of 'Britishness' or,

conversely, as its very essence, and “*We are all Immigrants*”, in which notions of ‘Britishness’ were explicitly linked to Britain’s history.

4.4.4.1 Britishness under threat. The content of a number of posts seemed to express fears that Britain might be ‘losing’ its ‘Britishness’ due to immigration. In particular, these fears tended to relate to posters’ perceived inability/restriction to express their ‘Britishness’ (i.e. not being allowed to “*fly the George Cross on St. George’s Day*”, or to “*refer to Christmas*”). In other words, here, Christian ‘Britishness’ is seen to be under threat. Along the same lines, some posters expressed a sense of identity threat by references to “*cultural genocide*” or “*cultural cleansing*” in relation to immigration. Here, such references create a sense of social identity threat (Branscombe et al., 1999; Reicher et al., 2008; Verkuyten, 2013) amongst those supporting Emma whilst simultaneously invoking conflict with the perceived out-group (i.e. foreigners).

Interestingly, most posters who self-identified as English appeared to endorse a British national identity (rather than an English one). In particular, respondents who explicitly referred to their English heritage rarely voiced concerns about losing their *Englishness* – instead, when national identity was seen to be under threat this tended to be expressed in terms of the *British* losing their ‘Britishness’. In addition, posters frequently differentiated between a British *we/this country* from a non-British *them/other countries* (Condor, 2000):

She has a valid point, albeit her language is really strong. The white, working classes in this country are becoming increasingly marginalised by the flood of immigrants from Africa, Asia and Eastern Europe. She's frustrated and angry. The present and preceding governments have failed her for years, and all she sees on public transport are ugly, threatening black and brown faces. Welcome to Tony 'War Criminal' Blair's vision of Britain for the 21st century. (Poster N)

Here, the British government is reproached for allowing “*the flood of immigrants*” into Britain – the poster thus distances him/herself from the government’s vision of Britain. This distancing is further emphasised by stating that the government “*have failed*” the “*White working classes*” which, together with references to “*ugly, threatening black and brown faces*”, underscore the poster’s sense of threat to his/her national identity. Also, the poster’s use of the term “*marginalisation*” is interesting here. First, the term conveys a sense of (relative) deprivation for which s/he holds immigrants responsible (see also above). Second, the use of this particular term here expresses a sense of identity threat by implying that White people have become a minority in Britain.

However, many other posters rejected this notion of *Britishness Under Threat* altogether. These posters argued in favour of immigration, suggesting that it enhances and enriches ‘Britishness’. This, in turn, was linked to calls for greater equal rights in terms of who should be allowed to call themselves British:

I'm British (white) and know I have no greater right to call myself British than those who became citizens today. I hope the ignorant minority of white nationalists will flow into our cultural past, but I doubt it. No society ever lives without ignorance, prejudice or racism completely. I would say I'm ashamed of this woman but I see no connection between myself and her just because of our race. Don't get frustrated by her ignorance, Britain's a multicultural society, and I'm proud of that. (Poster O)

Here, then, multiculturalism becomes the very essence of what it means to be British – which is in direct opposition to the views and behaviour depicted in the YouTube video. The poster clearly dis-identifies with Emma (i.e. “*see no connection with her*”) and distances him/herself from White nationalists in general (i.e. by referring to them as

"*ignorant*"). Rather than threatening British national identity, here multiculturalism is construed as defining 'Britishness'.

Other posters argued that throughout British history different cultural groups came to Britain and therefore shaped 'Britishness'. This not only gives people the right to call themselves British, it also undermines 'Britishness' as a fixed category. Some posters even suggested that "*We (i.e. the British) are all Immigrants*" – a claim which will be considered next.

4.4.4.2 "We are all immigrants". Many YouTube comments explicitly and positively embraced immigration, especially given Britain's colonial history. For example, some posters responded to those who rejected immigration by suggesting that, under British Empire rule, people from British colonies had not only been encouraged but, at times, forced to live in Britain:

Did anybody force the British to buy African slaves? By buying Africans, they made them British. The same history told me that the British used extreme force to capture several African settlements. They even sent the leaders of some of these settlements into exile. An example is King JaJa of Opobo. The British traveled for Months to look for Africa. They found Africa and raped her. They have to bear the consequences of their actions. No complaints is entertained from any British about Africans (Poster P)

This then raises the issue of who has the right to call him/herself British and bestow this identity on others. The specific argument here is that African descendants have the same right to call themselves British based on their ancestral history. Here, people who reject immigration are being reminded, that, although Britain's colonial history might appear glorious to some, British history is rooted in the (bleak) past of those living in the colonies – which, in turn, here justifies immigration.

The YouTube comments also challenged Emma's concept of 'Britishness' by referring to Britain's pre-colonial history. For example, some posters undermined Emma's notion of 'Britishness' by suggesting that:

(...) if you look at the history of britian [we] are pretty much all immigrants... a mix of roman and saxon and norman and viking... further back you have the celts and the pict's... who came from somewhere else in europe... (Poster Q)

Again, the argument here is that 'Britishness' is, in fact, a function of the nation's cultural diversity – and as this diversity is evident in ancient British history, no one should thus be considered more or less British. The implication here is that there is no inherent or fixed essence to 'Britishness'; instead, as Britain's history and present continue to be shaped by cultural diversity and multiculturalism, what it means to be British evolves and expands. This position, then, lead to an advocacy of tolerance and diversity, which is reflected in the following extract:

There are good and bad in every race. Every one has a right to there own opinion but no one has the right to disrespect and abuse people based on there race and religion this is not freedom of speech. White people are not the only ones who contribute to this country every where you go you will see black and brown doctors, teachers, politicians, police officers and soldier's many have died for this country. Do you know why? Because this is our country! We are different but equal! Peace (Poster R)

Overall, as a prevalent theme, *Contesting Britishness* encapsulated attempts to define 'Britishness' – in other words, it captured efforts to define 'us'. Specifically, respondents contested the criteria that determine who should and should not qualify as British, which often related to immigration. This contestation sometimes included perceived threats to posters' sense of Britishness (i.e. social identity threat; Branscombe et

al., 1999; Reicher et al., 2008). However, for many others, Britishness constituted a broad, flexible and evolving social category with immigration/multiculturalism at its very heart.

In sum, the current analysis indicates that YouTubers responded to explicit hate materials by attempting to define or contest the meaning of posters' national (i.e. here, predominantly British) identity. Responses also frequently served to position users according to their stance on racism. As such, whilst some YouTubers responded in a racist, aggressive and abusive manner, the majority of posters did not automatically endorse, or raise their responses beyond, the level of hatred displayed in the video clip.

4.5 Discussion

The objective of this study was to explore how general users of a particular SNS (i.e. YouTube) responded to online materials which could be construed as explicitly inciting hatred. To address this objective, comments in response to a YouTube video clip which depicted 'Emma', a White female train passenger racially abusing other passengers, were analysed.

Thematic Analysis (TA, Braun & Clarke, 2006) identified four superordinate themes: (1) *Making Sense of Emma*, (2) *Meeting Hatred with Hatred*, (3) *Us versus Them*, and (4) *Contesting Britishness*. *Making Sense of Emma* depicted attempts to explain Emma and her behaviour by invoking both intra-personal and social level explanations (i.e. alcohol/drug abuse, lack of intelligence, lack of educational and/or low socio-economic status) which also served to distance posters from Emma. *Meeting Hatred with Hatred* captured 'enflamed' responses to the video clip's content where hateful language was used, including calls for symbolic punishment, to express opposition to Emma and the views embodied by her. *Us versus Them* captured posters' attempts to construct in- and out-groups in order to position themselves according to their own stance on racism. This positioning frequently occurred with reference to immigration and related to expressions of

a sense of injustice and perceived relative deprivation. The final theme, *Contesting Britishness*, depicted attempts to define (or contest) the British national identity, especially in relation to Britain's colonial history and immigration. Whilst some YouTube posters expressed a sense of their national identity being under threat, others suggested that there is no such thing as 'Britishness' – and, for some, immigration/ multiculturalism constituted the very essence of 'Britishness'. These themes link to existing theories as well as previous empirical findings which will be discussed below.

Posters frequently attempted to distance themselves from Emma's behaviour by positioning her as an out-group member (i.e. a racist). This, in turn, was achieved through invoking negative stereotypes (i.e. low intelligence, low socio-economic and single mother status) and the use of hateful, dehumanising language. Similar distancing strategies have been observed in previous work (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Douglas et al., 2005; Erjavec & Kovačič, 2012; Faulkner & Bliuc, 2016; Moghaddam, 2005; Mols & Jetten, 2014; Steinfeldt et al., 2010) where the use of (hateful) language has been described as a means of creating negative social comparisons and reinforcing existing stereotypes – or 'othering'. Here, this 'othering' strategy served not only as a means of *Making sense of Emma* but, paradoxically, also led to *Meeting hatred with hatred*, as expressed in Emma's dehumanisation.

The dehumanisation of perceived out-group members has been observed previously as a common linguistic 'othering' strategy (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Faulkner & Bliuc, 2016; Moghaddam, 2005). Dehumanisation of perceived out-group members has also been associated with attempts to foster a sense of moral righteousness or moral disengagement (Cammaerts, 2009; Faulkner & Bliuc, 2016) and the creation of a sense of in-group superiority (Bliuc, McGarty, Hartley & Muntele Hendres, 2012; Mumford et al., 2008). For instance, Faulkner and Bliuc (2016) found that racists

used moral disengagement strategies such as dehumanisation to justify expressions of their views. Interestingly, in this study, dehumanisation was used by both those who supported or opposed racist attitudes to justify expressing their views. In line with this, some sub-themes (i.e. *Sense of deprivation*, *Sense of injustice* and *Britishness under threat*) echo some of Faulkner and Bliuc's (2016) moral disengagement themes.

Furthermore, one of the functions of dehumanisation may be to enable in-group members to justify punitive treatments of individual out-group members and out-groups (Moghaddam, 2005) – which, in this study, was reflected in posters' demands for Emma's symbolic punishment. Similar calls for symbolic punishment have also been observed in other online hate contexts (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Doughty et al., 2014; Erjavec & Kovačič, 2012). Some posters' attempts to counter Emma's apparent racism with further expressions of hatred (i.e. *Meeting hatred with hatred*) seem to be in line with existing 'flaming' research (Coffey & Woolworth, 2004; Doughty et al., 2014, Lange, 2007; Moor et al., 2010). In particular, there was some evidence of 'flaming' (i.e. use of aggressive/abusive/racist language to oppose certain web content) in the online comments examined here. Specifically, posters appeared to use hateful language, or 'flames', not only to express their disagreement with the video clip, but also in order to mirror the norms conveyed in the video clip (i.e. as evidenced by Emma's own aggressive, racist, insulting and hostile language and behaviour). This corresponds to previous research on online aggression which suggests that aggressive expressions on the Internet are often a reaction to perceived aggression from others (Law, Shapka, Domene & Gagne, 2012).

The most prominent themes identified in the current analysis – namely, *Us versus Them* and *Contesting Britishness* – explicitly tapped into social identity processes and, as such, not only correspond to previous findings but also fit within a broader social identity approach (Blackwood et al., 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979;

Turner et al., 1987; Verkuyten, 2013). For example, Angie et al. (2011) related the use of negative stereotyping and categorical thinking – as evidenced by ‘us’ versus ‘them’ – to social identity formation. In particular, Angie et al. (2011) found the use of categorical thinking more prominent in non-violent ideological groups than in violent ones.

Categorical ‘us-versus-them’ thinking was also observed by Blackwood et al. (2012), Bliuc et al. (2012) and Reicher et al. (2008). In line with previous work, in order to confirm their identification with like-minded respondents, or their group identity (i.e. here, non-racists), many posters in this study adopted similar categorical thinking strategies to those identified in previous work (e.g. Angie et al., 2011; Charteris-Black, 2006).

Posters who supported Emma’s behaviour also distinguished between ‘us’ and ‘them’ – specifically by using terminology that depicted ‘us’ as good and ‘them’ as bad. This, too, corresponds to previous observations that positive self-representation is often pitched against negative other-presentation (Douglas et al., 2005; Verkuyten, 2013) in order to create negative social comparisons which, in turn, aims to establish a sense of legitimacy and to persuade like-minded others (Charteris-Black, 2006). Here, negative other-presentations was frequently linked to expressions of a sense of injustice or relative deprivation. For example, posters who endorsed Emma’s views tended to claim that immigrants and ethnic minorities were in receipt of preferential treatment by the State or by the police - which, in turn, mirrors previous findings (e.g. Coffey & Woolworth, 2004; Douglas et al., 2005; Mols & Jetten, 2014). These posters also invoked the notion of socio-economic insecurity as a consequence of immigration; attributing blame to immigrants has previously been linked to scapegoating, disgruntled sentiments, prejudice and hate crimes towards minority groups (Bowling & Phillips, 2003; Sibbitt, 1997; Young, 1999; for a

detailed review see Walters, 2010).³

Taken together, these findings closely relate to the concept of social identity threat (Branscombe et al., 1999; Reicher et al., 2008, Verkuyten, 2013) which derives from the social identity perspective (e.g. Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner's, 1979; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). Whilst there are different types of social identity threat, the current findings relate most closely to what Branscombe et al. (1999) have termed 'distinctiveness threat' (i.e. the fear that one's group distinctiveness is prevented or undermined). This suggests that perceived threats to one's group's identity may lead to discrimination including expressions of overt racism and hostility (Perreira, Vala & Costa-Lopes, 2010; Reicher et al., 2008; Verkuyten, 2013).

In line with previous findings (e.g. Faulkner & Bliuc, 2016), posters in this study, who supported racist views, attempted to justify their opinions by expressing a perceived sense of threat to the distinctiveness of their social (i.e. national) identity (Branscombe et al., 1999) on the basis of immigration. It has been shown previously that one of the responses to multiculturalism can be the expression of a perceived (national) identity threat (Mols & Jetten, 2014; Sindic; 2008). Issues relating to perceived social identity threat were also evident in the final theme – namely, *Contesting Britishness*. In particular, posters who supported racist views invoked a sense of socio-economic threat to the White British working class.

Another way in which the themes *Us versus Them* and *Contesting Britishness* tapped into social identity processes relates to YouTubers' contestation of their national

³ At the time of the video clip's upload (e.g. October 2011), Britain was in an economic recession and characterised by socio-economic insecurity. This may therefore have exacerbated posters' sense of relative material and financial deprivation.

identity as British. In particular, Faulkner and Bliuc (2016) previously observed how Australians disputed in online newspaper comments the perceived values and characteristics considered representative of their national identity. In addition, differences in people's versions of 'Britishness' have been identified previously (e.g. Condor, 1997). In particular, differing accounts of 'Britishness' are frequently found in political discourse where they serve as rhetorical tools to influence people's interests and behaviours according to political agendas (Sindic, 2008). Therefore, as political agendas change, so might the boundaries/definitions of 'Britishness'.

Along similar lines, national identities are not fixed and, as Cohen (1994) has argued, their boundaries are generally 'fuzzy'. Thus, national identities are continuously re-defined through interactions with different 'others' (Condor, 2000; Hopkins & Murdoch, 1999).

4.5.1 Exposure to Online Hate Materials and Escalation

Coffey and Woolworth (2004) observed that some online discussions can be characterised by escalation (i.e. in terms of racism, aggression, abuse). However, in the context of the present study, the author did not find any straightforward evidence of escalation; instead, posters' responses seemed more complex and, as discussed above, often contested relevant social categories (i.e. 'us' versus 'them', 'Britishness'). One reason for the lack of evidence of escalation here is likely to be found in the relative short time-span (i.e. 48 hours) posters had available to respond to the video clip (i.e. before the clips' removal from YouTube).

Another likely reason is the asynchronous nature of the responses under investigation here and the corresponding non-sequential way in which posts were displayed on YouTube: unlike in Coffey and Woolworth's study (2004), given the volume of simultaneous responses to the YouTube video, posters were rarely able to respond

sequentially (i.e. directly) to specific others. Some respondents attempted to overcome this issue by creating a response sequence addressing other posters directly (i.e. starting their responses with an '@' and indicating the recipient's username). Yet, by the time their posts were displayed, given the sheer number and speed of postings, these attempts at creating response sequences would have been hard to identify visually (i.e. they would have appeared 'lost' amongst hundreds of other comments).

The potential differences between synchronous versus asynchronous communication and user density in online spaces have been highlighted previously (Wojcieszak, 2010). In particular, Wojcieszak (2010) questioned whether online response synchrony and user density would facilitate or prevent people from responding in a more polarised (i.e. escalated) manner over time. Thus, the relative short time, asynchrony and density of the comments analysed here might, in part, have prevented further polarisation and escalation of hateful responses. The specific relationship between response time, asynchrony, density and escalation of hateful responses therefore warrants further investigation in order to clarify the possible factors contributing to polarised responses.

Finally, whilst collating the online comments, it was observed that several respondents were removing their comments from the website. This removal seemed prompted by other respondents' appeals to the service providers (i.e. YouTube) or the police to delete or investigate certain comments for their perceived offensive and illegal nature (i.e. hate speech and/or incitement to racial hatred). Therefore, the removal of posts seemed related to posters' concerns about potential legal consequences and a decreased sense of online anonymity. More importantly, the observed removal of posts suggests that a certain degree of self-policing is practiced online.

It was also observed that some posters blocked or spammed others who expressed racist views, as well as advised others to do the same (via commenting). Consequently,

responses by blocked or spammed posters became no longer visible to anyone on YouTube, suggesting that online hate materials can, in fact, elicit self-censorship. This self-censorship, in turn, seems to further support the finding that social networking users do not always automatically endorse online hate materials.

4.5.2 Limitations

First, as with qualitative work in general, one of the limitations of this study lies in the fact that the current findings cannot be generalised (e.g. across all online hate materials or different SNSs/SMPs). Here, responses to a certain, racially motivated hate incident on one specific SNS/SMP (i.e. YouTube) were explored. The author therefore cannot claim to explain hatred across all SNSs/SMPs or all hate-motivated incidents, such as those based on disability, faith, gender, transgender or subculture. Instead, the broad focus of the current research question aimed to provide a starting point for research by exploring the nature of online replies to explicit hate materials. In particular, the present study aimed to shift the currently dominant ‘end-product approach’ (i.e. already polarised individuals/groups and their content/interactions) to a focus on the potential processes underlying the development of polarisation in the context of online hatred. It was found that whilst online racism can lead to some degree of polarisation, the majority of responses did not suggest polarisation per se, but instead indicated social identity processes at play (i.e. reflecting the construction of in- and out-groups, illustrating respondents’ group affiliations).

Second, whilst one of the study’s strengths lies in the realistic nature of the data, this observational approach prevents the author from making inferences about the representativeness of the responses (e.g. the sample might not represent users across other SNSs/SMPs). The author neither had access to respondents’ demographics, nor could she determine whether respondents posted with multiple or different user names (i.e.

identities). Yet, given the substantial number of comments in response to the video clip (i.e. over 71,000 posts), it is unlikely that these were exclusive to a specific (demographic) group – instead, there was evidence of some diversity in backgrounds (e.g. gender, ethnicity, geographic location).

Finally, as in previous studies (e.g. Coffey & Woolworth, 2004), the findings do not allow the author to draw inferences about the effects that the relatively short period of time in which people were able to respond to the video clip might have had on the content of their posts. In particular, it remains unclear whether longer-term online interactions result in repeated and/or increased expressions of hatred or whether long-term interactions/discussions change direction and automatically lose focus, therefore reducing or diverting expressions of hatred.

4.5.3 Conclusion

Expressions of hatred (e.g. racism) are more and more becoming a feature of life online, especially on Social Networking Sites and Social Media Platforms (e.g. YouTube). It is therefore becoming increasingly important to understand how ‘everyday’ online users respond to and engage with such hateful materials. Many YouTubers who responded to the hateful online material analysed here condemned its content. As such, the findings obtained in the current context indicate that online hatred may not always be automatically endorsed by online users; instead, a degree of contestations in online users’ responses was observed. Whilst the current findings cannot be readily generalised to all other forms of social networking and/or forms of hatred, they nevertheless show how YouTubers engaged with racial hatred and each other, how quickly they engaged and how willing they were to engage. In addition, the sheer weight of the number of responses shows that this racial incident resonates within a section of society. The author therefore hopes that these findings can also provide a springboard for future research which should take into account

not only a wider variety of hate materials but also different Social Networking Sites/Social Media Platforms.

4.6 References

- Agarib, A. (2016, January 1). Man arrested for online hate crime. *Khaleej Times*. Retrieved from <http://www.khaleejtimes.com/article/20160101/ARTICLE/301019214>
- Allen, M., Bromley, A., Kuyken, W., & Sonnenberg, S. (2009). Participants' experiences of mindfulness-based cognitive therapy: "It changed me in about every way possible". *Behavioural and Cognitive Psychotherapy*, *37*, 413-430. DOI: 10.1017/S135246580999004X
- Anti-Defamation League (2005). *Attitudes toward Jews in twelve European countries* (May 2005). Retrieved from www.adl.org: http://www.adl.org/israel/Eur_Poll_Israel_May_2005.pdf
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., Hoang, T. S., Bernard, D. R., Hughes, M. G., Connelly, S., O'Hair, H. D., & Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology*, *41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Baron, N. S. (2004). See you online: Gender issues in college students use of instant messaging. *Journal of Language and Social Psychology*, *23*, 397-423. DOI:10.1177/0261927X04269585
- BBC News (2012, February 10). *Derby men jailed for giving out gay death call leaflets*. *BBC News*. Retrieved from <http://www.bbc.co.uk/news/uk-england-derbyshire-16985147>
- BBC News (2016, January 12). Teenager given probation for racist slogan. *BBC News*. Retrieved from <http://www.bbc.com/news/uk-northern-ireland-35294697>

- Blackwood, L., Hopkins, N., & Reicher, S. (2012). Divided by a common language? Conceptualising identity, discrimination and alienation. In K. J. Jonas & T. A. Morton (Eds.), *Restoring civil societies: The psychology of intervention and engagement following crisis* (222- 236). Chichester: John Wiley & Sons Inc.
- Blackwood, L., Hopkins, N., & Reicher, S. (2013). I know who I am, but who do they think I am? Muslim perspectives on encounters with airport authorities. *Ethnic and Racial Studies*, *36*, 1090-1108. DOI: <http://dx.doi.org/10.1080/01419870.2011.645845>
- Blackwood, L., Hopkins, N., & Reicher, S. (2015). Turning the gaze on 'us': The role of authorities in the alienation of minorities. *European Psychologist*, *18*, 245-252. DOI: 10.1027/1016-9040/a000151
- Blazak, R. (2001). White boys to terrorist men: Target recruitment of Nazi skinheads. *American Behavioral Scientist*, *44*, 982–1000. DOI: 10.1177/00027640121956629
- Bliuc, A. M., McGarty, C., Hartley, L. K., & Muntele Hendres, D. (2012). Manipulating national identity: the strategic use of rhetoric by supporters and opponents of the 'Cronulla riots' in Australia. *Ethnic and Racial Studies*, *35*, 2174-2194. DOI: 10.1080/01419870.2011.600768
- Bostdorff, D. M. (2004). The internet rhetoric of the Ku Klux Klan: A case study in website community building run amok. *Communication Studies*, *55*, 340-361. DOI: 10.1080/10510970409388623
- Bowling, B., & Phillips, C. (2003). Racist victimization in England and Wales. In D. Hawkins (Ed.), *Violent crime, assessing race & ethnic difference* (pp. 154-170). Cambridge: Cambridge University Press.

- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The content and context of social identity threat. In R. Spears, B. Doosje & N. Ellemers (Eds.), *Social identity: Context, commitment, content* (pp. 35-55). Oxford: Blackwell.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. DOI: 10.1191/1478088706qp063oa
- British Psychological Society (2013). *Ethics guidelines for internet-mediated research*. Retrieved from <http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf>
- Burrows, T. (2016, March 23). Talent agency boss who ‘confronted a Muslim woman in London and asked her to explain Brussels’ is arrested for ‘inciting racial hatred’. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/news/article-3506275/publicity-boss-asked-Muslim-woman-street-explain-Brussels-deluged-angry-hilarious-Tweets-reply.html>
- Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces: The case of north Belgian extreme right discourses. *International Journal of Cultural Studies*, 12, 555-575. DOI: 10.1177/1367877909342479
- Castle, C. (2013, October 27). YouTube awards tarnished by YouTube hate videos and jihadi recruitment. *Music, Technology, Policy*. Retrieved from <https://musictechpolicy.wordpress.com/2013/10/27/youtube-awards-tarnished-by-youtube-hate-videos-and-jihadi-recruitment/>
- Charteris-Black, J. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17, 563-581.
DOI: 10.1177/0957926506066345

- Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies*, 65, 57-70. DOI: 10.1016/j.ijhcs.2006.08.009
- Citron, D. K., & Norton, H. L. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review*, 91, 1435-1484, University of Maryland Legal Studies Research Paper No. 2011-16.
- Coffey, B., & Woolworth, S. (2004). "Destroy the scum, and then neuter their families:" The web forum as a vehicle for community discourse? *The Social Science Journal*, 41(1), 1-14. DOI: 10.1016/j.soscij.2003.10.001
- Cohen, R. (1994). *Frontiers of Identity: The British and others*. London: Longman.
- Condor, S. (1997). Having history: A social psychological exploration of Anglo-British stereotypes. In C. C. Barfoot (Ed.), *Beyond pug's tour: National and ethnic stereotyping in theory and literature practice*, (pp. 213-253). Amsterdam: Rodopi.
- Condor, S. (2000). 'Pride and prejudice': Identity management in English people's talk about "this country". *Discourse & Society*, 11, 175-205. DOI: 10.1177/0957926500011002003
- Crown Prosecution Service (2010, February 17). Teenager admits incitement to racial hatred over YouTube videos. *CPS News*. Retrieved from http://www.cps.gov.uk/news/latest_news/teenager_admits_incitement_to_racial_hatred/
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354. DOI: 10.1037/h0047358

- DeKoster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me'. *Information, Communication & Society, 11*, 1155-1176. DOI: 10.1080/13691180802266665
- Doane, A. (2006). What is racism? Racial discourse and racial politics. *Critical Sociology, 32*, 255-274. DOI: 10.1163/156916306777835303
- Doughty, M., Lawson, S., Linehan, C., Rowland, D., & Bennett, L. (2014). Disinhibited abuse of othered communities by second-screening audiences. *TVX, 14*, 55-62. DOI: 10.1145/2602299.2602311
- Douglas, K. (2007). Psychology, discrimination and hate groups online. In A. Joinson, K. McKenna, T. Postmes & U. D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 155–163). New York: Oxford University Press.
- Douglas, K. M., McGarty, C., Bliuc, A. M., & Lala, G. (2005). Understanding cyberhate: Social competition and group creativity in online white supremacy groups. *Social Science Computer Review, 23*, 68-76. DOI: 10.1177/0894439304271538
- Erjavec, K., & Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites'. *Comments, Mass Communication and Society, 15*, 899-920. DOI: 10.1080/15205436.2011.619679
- ETHNOS (2005). *Citizenship and belonging: What is Britishness?* Commission for Racial Equality (CRE). London: Ethnos Research and Consultancy.
- Ezekiel, R. S. (2002). An ethnographer looks at neo-Nazi and Klan groups: The racist mind revisited. *American Behavioral Scientist, 46*, 51-71. DOI: 10.1177/0002764202046001005
- Faulkner, N., & Bliuc, A. M. (2016). "It's okay to be racist": Moral disengagement in online discussions of racist incidents in Australia. *Ethnic and Racial Studies, 39*, 2545-2563. DOI: 10.1080/01419870.2016.1171370.

- Foxman, A. H., & Wolf, C. (2013). *Viral hate containing its spread on the internet*. New York: Palgrave Macmillan.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400. DOI: 10.1016/0191-8869(86)90014-0
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193. DOI: 10.1111/j.1745-6916.2007.00036.x
- Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, 3, 29-44. DOI: 10.1111/j.1530-2415.2003.00013.x
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58, 177-193. DOI: 10.1111/1540-4560.00255
- Gleibs, I. H., Sonnenberg, S., & Haslam, C. (2014). We get to decide”: The role of collective engagement in counteracting feelings of confinement and lack of autonomy in residential care. *Activities, adaptation & aging*, 38, 259-280. DOI:10.1080/01924788.2014.966542
- Green, D. P., Abelson, R. P., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice, & D. T. Miller (Eds.), *Cultural Divides*, (pp. 429-464). New York: Russell Sage Foundation.
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6, 215-242. DOI: 10.1515/jplr.2010.011

- Herring, S., Job-Sluder, K., Scheckler, R., & Brab, S. (2002). Searching for safety online: “Trolling” in a feminist forum. *The Information Society: An International Journal*, 18, 371-384. DOI: 10.1080/01972240290108186
- Hopkins, N., & Murdoch, N. (1999) The role of the ‘other’ in national identity: Exploring the context-dependency of the national in-group stereotype. *Journal of Community and Applied Social Psychology*, 9, 321-338. DOI: 10.1002/(SICI)1099-1298(199909/10)9:5<321::AID-CASP510>3.0.CO;2-M
- Internet Live Stats (2016, 21 March). *Internet users*. Retrieved from <http://www.internetlivestats.com/internet-users/>
- Internet Live Stats (2016, 21 March). *Facebook users*. Retrieved from <http://www.internetlivestats.com/>
- Jacks, W., & Adler, J. R. (2015). A typology of online hate crime. *Open Access Journal of Forensic Psychology*, 7, 64-89. Retrieved from <http://eprints.mdx.ac.uk/18930/1/Jacks%20and%20Adler%202015.pdf>
- Jewish News Online (2015, December 17). Extremist Bonehill jailed for three years for inciting racial hatred. *Jewish News Online*. Retrieved from <http://jewishnews.timesofisrael.com/extremist-bonehill-guilty-of-inciting-racial-hatred/>
- Joinson, A. N. (1998). Causes and implication of disinhibited behaviour on the internet. In J. Gackenbach (Ed.), *Psychology and the internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43-60). San Diego: Academic Press.
- Kazakhstan Newslines (2015, August 3). New person sentenced for ‘inciting religious hatred’ in Kazakhstan. *Kazakhstan Newslines*. Retrieved from <http://www.newslines.kz/news.php?y=15&m=08&d=03#0>

- Lange, P. G. (2007). *Commenting on comments: Investigating responses to antagonism on YouTube*. Paper presented at the Annual Conference of the Society for Applied Anthropology, Tampa, Florida. Retrieved from <https://www.yumpu.com/en/document/view/14888058/commenting-on-comments-investigating-responses-to-antagonism->
- Law, D. M., Shapka, J. D., Domene, J. F., & Gagne, M. H. (2012). Are cyberbullies really bullies? An investigation of reactive and proactive online aggression. *Computer in Human Behavior, 28*, 664-672. DOI: 10.1016/j.chb.2011.11.013
- Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist, 45*, 927-957. DOI: 10.1177/0002764202045006003
- Levin, B. (2002). Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioral Scientist, 45*, 958-988. DOI: 10.1177/0002764202045006004
- Mail Online (2016, March 8). Australian behind Singapore hate site convicted of sedition. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/wires/afp/article-3482015/Australian-Singapore-hate-site-convicted-sedition.html>
- McNab, H. (2015, February 12). Crucifixions and a \$50 living allowance: Aussie who joined ISIS reveals what life is really like in Syria after defecting because he didn't like Westerners being executed. *Daily Mail Australia*. Retrieved from <http://www.dailymail.co.uk/news/article-2950469/It-s-not-like-YouTube-Former-ISIS-member-defected-terror-group-reveals-escaped-brutal-executions-Westerners-forced-out.html>

- McNamee, L. G., Peterson, B. L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs*, 77, 257-280. DOI: 10.1080/03637751003758227
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist*, 60, 161-169. DOI: 10.1037/0003-066X.60.2.161
- Mols, F., & Jetten, J. (2014). No guts, no glory: How framing the collective past paves the way for anti-immigrant sentiments. *International Journal of Intercultural Relations* 43, 74-86. DOI: 10.1016/j.ijintrel.2014.08.014
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26, 1536-1546. DOI: 10.1016/j.chb.2010.05.023
- Mumford, M. D., Bedell-Avers, K. E., Hunter, S. T., Espejo, J., Eubanks, D., & Connelly, M. S. (2008). Violence in ideological and non-ideological groups: A quantitative analysis of qualitative data. *Journal of Applied Social Psychology*, 38, 1521-1561. DOI: 10.1111/j.1559-1816.2008.00358.x
- Nirvana News (2016, January 5). Trump used in terrorist recruitment video. *Nirvana News*. Retrieved from <https://www.youtube.com/watch?v=cv1aU3pG3BQ>
- O'Sullivan, P. B., & Flanagan, A. (2003). Reconceptualizing "flaming" and other problematic communication. *New Media and Society* 5, 69-94. DOI: 10.1177/1461444803005001908
- Parisi, P. (1998). A sort of compassion: The Washington post explains the crisis in urban America. *Howard Journal of Communications*, 9, 187-203. DOI: 10.1080/106461798246989

- Peddinti, S., Ross, K., & Cappos, J. (2014). On the internet, nobody knows you're a dog: A twitter case study of anonymity in social networks. *Proceedings from COSN '14: The second ACM conference on Online social networks*. Retrieved from <http://cosn.acm.org/2014/files/cosn025f-peddintiA.pdf>
- Pereira, C., Vala, J., & Costa-Lopes, R. (2010). From prejudice to discrimination: The legitimizing role of perceived threat in discrimination against immigrants. *European Journal of Social Psychology, 40*, 1231–1250. DOI: 10.1002/ejsp.718
- Perry, B. (2000). “Button-down terror”: The metamorphosis of the hate movement. *Sociological Focus, 33*, 113–131. DOI: 10.1080/00380237.2000.10571161
- Phillips, T. (2015, December 22). Pu Zhiqiang given three-year suspended sentence. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2015/dec/22/pu-zhiqiang-chinese-human-rights-lawyer-sentenced-to-three-years>
- Poch, R. (2015, November 15). Fighting incitement on social media, “they can’t” gets 40,000 hate videos removed from YouTube. *Breaking Israel News*. Retrieved from: <http://www.breakingisraelnews.com/53013/they-cant-launches-campaign-against-palestinian-incitement-social-media-terror-watch/#toBExsVzYqfs3VLg.97>
- Radio Free Europe Radio Liberty (2015, July 31). Another Kazakh sentenced for inciting ethnic hatred online. *Radio Free Europe Radio Liberty*. Retrieved from <http://www.rferl.org/a/kazakhstan-ethnic-hatred-man-sentenced/27163239.html>
- Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior, 29*, 610-615. DOI: 10.1016/j.chb.2012.11.011

- Reicher, S., Haslam, S. A., & Rath, R. (2008). Making a virtue of evil: A five-step social identity model of the development of collective hate. *Social and Personality Psychology Compass*, 2, 1313–1344. DOI:10.1111/j.1751-9004.2008.00113.x
- Reuters (2015, December 30). Russian court jails blogger for five years for ‘extremist’ posts. *Reuters*. Retrieved from <http://uk.mobile.reuters.com/article/idUKKBN0UD16O20151230>
- Risen, T. (2014). Twitter, Facebook, YouTube grapple with Islamic State censorship. *US News*. Retrieved from <http://www.usnews.com/news/articles/2014/09/05/twitter-facebook-youtube-navigate-islamic-state-censorship>
- Robi, E. (2013). What computer forensics tell us about the Boston bombers. *Elluma Discovery*. Retrieved from <http://www.elluma.com/what-computer-forensics-tells-us-about-the-boston-bombers/>
- Runciman, W. G. (1966). *Relative deprivation and social justice*. London: Routledge.
- Sibbitt, R. (1997). *The perpetrators of racial harassment and racial violence*. Home Office Research Study 176. London: Home Office.
- Sindic, D. (2008). *National identities: Are they declining?* Futurelab: UK Department for Children, Schools and Families’ Beyond Current Horizons Project. Retrieved from http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/final_sindic_nationalidentities_20081201.pdf
- Spencer, R. (2015, December 31). Kazakhstan: Convert from Islam to Christianity gets 2 years’ prison for inciting religious hatred. *Jihadwatch.org*. Retrieved from <https://www.jihadwatch.org/2015/12/kazakhstan-convert-from-islam-to-christianity-gets-2-years-prison-for-inciting-religious-hatred>

- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano Jr., L. A., Benton, E., & Steinfeldt M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology, 16*, 362-371. DOI: 10.1037/a0018692
- Stern, K. S. (2002). *Hate and the internet*. New York: The American Jewish Committee.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*, 321-326. DOI: 10.1089/1094931041291295.
- Swinford, S. (2013). Woolwich attack: Google must block hate preachers, says Theresa May. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/10081617/Woolwich-attack-Google-must-block-hate-preachers-says-Theresa-May.html>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Pacific Grove, CA: Books/Cole.
- Thomas, S. (1998). Race, gender, and welfare reform: The antinationalist response. *Journal of Black Studies, 28*, 419-446.
- Turnage, A. K. (2008). Email flaming behaviors and organizational conflict. *Journal of Computer-Mediated Communication, 13*, 43-59. DOI: 10.1111/j.1083-6101.2007.00385.x
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A Self-Categorization Theory*. Oxford & New York: Blackwell.

- Turpin-Petrosino, C. (2002). Hateful sirens . . . who hears their song? An examination of student attitudes toward hate groups and affiliation potential. *Journal of Social Issues*, 58, 281–301. DOI: 10.1111/1540-4560.00261
- Verkuyten, M. (2013). Justifying discrimination against Muslim immigrants: Out-group ideology and the five-step social identity model. *British Journal of Social Psychology*, 52, 345-360. DOI: 10.1111/j.2044-8309.2011.02081.x
- Walters, M. A. (2011). A general theories of hate crime? Strain, doing difference and self control. *Critical Criminology* 19, 313-330. DOI: 10.1007/s10612-010-9128-2
- Wojcieszak, M. (2010). ‘Don’t talk to me’: Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media Society* 12, 637-655. DOI: 10.1177/1461444809342775
- Young, J. (1999). *The exclusive society: Social exclusion, crime and difference in late modernity*. London: Sage Publications.
- YouTube (n.d.). *YouTube-statistics*. Retrieved from: <https://www.youtube.com/yt/press/en-GB/statistics.html>
- Zhou, Y., Reid, E., Qin, J., H., Chen, H., & Lai, G. (2005). US domestic extremist groups on the web: Link and content analysis. *IEEE Intelligent Systems*, 20, 44–51. DOI: 10.1109/MIS.2005.96

Chapter 5

The Permissibility of Online Hatred: Exploring the Relationships between Hatefulness and Criminality, Online and Offline Hatred and the Role of Anonymity in Expressing Hatred

5.1 Abstract

Although it has been argued that hate on the Internet is increasing, there is still a lack of clarity regarding people's attitudes towards the permissibility of expressions of online hatred. In this study, 164 participants completed an online survey, which explored three areas of interest concerning online hatred: (1) what constitutes online hatred? (2) the relationship between online and offline hatred (3) the role of online anonymity in expressing online hatred. The results revealed that the evidence concerning the permissibility of online hatred is mixed. In particular, on the one hand, participants did not associate online with offline hatred, objected to legal consequences for inciting online hatred, and blamed the victims for their online abuse rather than the posters of hateful web content. On the other hand, participants largely agreed on which content classified as hateful and which content should be considered criminal and admitted that online anonymity aids their own anti-normative behaviour, despite most of them not actually feeling anonymous online. Moreover, results revealed that there is no consensus on whether the consequences of online or offline hatred are worse for its victims. Overall, the mixed results concerning the permissibility of online hatred highlight the complexities that online hatred presents for those at the brunt of it.

5.2 Introduction

5.2.1 Background

Hateful online content has increased in recent years (European Commission, 2016; Heller, as cited by North, 2016; Jacks & Adler, 2015), with some reports claiming that, in 2015, online hate speech spiralled out of control (Wolfram, 2016). For example, according to Facebook, the number of general user violations amounts to one million a day, compared to Twitter's 480,000 racist tweets per month (Wolfram, 2016). There has been a marked increase over time in the frequency of news headlines on hateful online content, expressions of online hatred and their consequences (i.e. hate offences) as well as in the number of people witnessing online hate crimes (BBC News, 2016; Graham, 2016).

Consequently, Jourová, the European Commissioner for Justice, Consumer and Gender Equality (cited by the European Commission, 2015) and others (e.g. ACPO, 2014; Angie et al., 2010; Foxman & Wolf, 2013; Home Office, 2009; OSCE, 2010) have expressed concern that “*there is growing evidence that online incitement to hatred leads to violence offline*”. Jourová (as cited by the European Commission, 2016) argued that online hate speech not only distresses the people it targets, but the fear of being targeted by online hate speech can also keep people away from Social Networking Sites (SNSs) and Social Media Platforms (SMPs), and thus it limits the digital space for freedom of expression. Supporting this, the continuous growth of online hate speech has led to the decision by some news websites (e.g. Bloomberg, Chicago Sun Times, Recode, Reuters) to remove (temporarily) their online commenting features (Wolfram, 2016).

According to police chiefs in the UK, online hate materials, and particularly online anonymity, play an increasing role in individuals and communities being targeted online on the grounds of their race, religion, disability, gender or transgender (Oryszczuk, 2016). In addition, Sukhni (as cited by Laubner, 2016) argued that Social Networking Sites/Social

Media Platforms (SNSs/SMPs; e.g. Facebook, Twitter, Instagram, YouTube) play an essential role in people's self-radicalisation. Sukhni based his argument on the fact that hateful online content (e.g. emotionally charged films, ideological content, instructions to build bombs/commit acts of terror) is professionally tailored and displayed to its intended target groups. Such concerns contribute towards the overall perception by governments, academics and Non-Governmental Organisations (NGOs) that hate on the Internet can be dangerous.

Governments across the world have, in turn, begun to amend and toughen existing legislation, introducing new laws to govern the online incitement to hatred. One such strategy includes the new European Union Internet Forum, a voluntary partnership between Government, Europol and major Internet companies, which was founded in December 2015. The forum aims to detect and address harmful online materials (European Commission, 2015). For example, the forum introduced the so called 'Code of Conduct' to tackle online hate across the European Union, to which Facebook, Google, Microsoft, Twitter and YouTube, signed up in May 2016 (Burlecu, 2016). Since then, some Internet companies have already started to set aside resources to introduce new tools to tackle online hate speech (Burlacu, 2016). For example, Microsoft updated their terms and services, introducing web forms enabling people to report hateful online content (Pradeep, 2016). In addition, Google Chrome recently introduced a new plugin called 'Hate Free' which detects and flags up hateful language to all online users accessing content which is considered hateful (Tech City News, 2016).

In addition, changes in legislation governing the incitement of online hatred have led to global increases in the number of legal cases against individuals conducting such behaviour (e.g. Creighton, see: Taylor, 2016; Choudary & Rahman, see: Crown Prosecution Service, 2016; Jumaa, see Jewish Telegraphic Agency, 2016; Takagi &

Kaiheng, see: Mail Online, 2016; Wilders, see: Mail Online, 2016; Alkhanashvili, Kazakhstan Newswire, 2015; Berlin woman, see: Berliner Zeitung, 2015; Bonehill-Paine, see: Jewish News Online, 2015; Yklas Kabduakasov, see: Spencer, 2015; Mukhtar, see: Radio Free Europe Radio Liberty, 2015; Pu Zhiquaing, see: Phillips, 2015; Shevtsova-Valova, see: Radio Free Europe Radio Liberty, 2015; Topham, see: CBC News, 2015; Tyumentsev, see: Reuters, 2015). For instance, in the last five years, arrests in London for expressions of hatred on SNSs have risen by 37% (Corfield, 2016). Yet, despite evidence of the prevalence and dangers of online hate (e.g. victimisation, recruitment to online hate groups resulting in radicalisation, social division, mobilisation towards acts of hatred), it remains largely unknown how permissible expressions of online hate are perceived to be by those who frequently come in contact with it (i.e. ‘everyday’ online/SNS/SMP users). For example, despite existing legislation, it remains largely unknown if SNS/SMP users agree on what web content constitutes online hatred, whether such content should be criminalised or not, whether online users accept the consequences of online victimisation and whether online anonymity still (i.e. despite technical and legal advancements) plays a role in negative online behaviour. These specific aspects of the permissibility of online hatred will therefore be considered in more detail next.

5.2.2 What Constitutes Online Hate?

As already indicated, so far, there is little evidence regarding people’s views on what, exactly, constitutes online hate and to what extent expressions of online hatred are regarded as permissible. Specifically, there seems to be a lack of what (kind of content) constitutes online hate and to what extent it is endorsed or rejected (Rohlfing & Sonnenberg, 2016). In addition, the results of a recent survey carried out by Avon and Somerset constabulary exploring how Twitter users would respond to witnessing online hate crimes, indicate that 70% regarded online hate crimes as banter (Churchill, 2016).

Moreover, only 11% of Twitter users indicated that they would report online hate crimes to the police (Churchill, 2016). Along similar lines, Moor, Heuvelman and Verleur (2010) found that YouTubers perceived flaming – i.e. “*displays of hostility by insulting, swearing or through use of otherwise offensive language*” (Moor, 2007) – as a ‘side-effect’ of respecting freedom of speech and not as a real or potential offline problem for themselves. However, Moor et al. (2010) also found that most YouTubers acknowledged that flaming is a problem for a minority of other YouTubers.

Finally, the extent to which people’s perceptions concerning expressions of online hate match legal definitions of online hate crimes also remains unknown. For example, Rohlfiing and Sonnenberg (2016) noted that whilst some YouTubers identified expressions against other people’s national identity as hateful, racist and criminal, others branded them as patriotic and called upon their right to freedom of expression. Therefore, people’s perceptions of what constitutes hatefulness, and what is regarded as criminal, seem to differ amongst online users.

5.2.3 Online versus Offline Hatred

As already mentioned, it has been argued that expressions of online hate can lead to hateful acts offline (e.g. Jourová, 2015; 2016). Yet, as with the lack of consensus about what constitutes online hate and about the relationship between hatefulness and criminality, the extent to which offline hateful acts are regarded as a consequence of online expressions of hate remains unclear. For example, and although not representing the specific focus or the main findings of their study, Rohlfiing and Sonnenberg (2016) observed that YouTubers’ attitudes towards the consequences of online hate varied (i.e. some responses implied that they believed that online hate can lead to offline hate, whereas others implied that online hatred is harmless online communication).

If online users disregard the consequences of online victimisation through expressions of hatred, they might consider such expressions to be acceptable and risk-free – which, in turn, would counteract some of the aims governments and other organisations have tried to achieve with the new strategies mentioned above. In addition, strategies to prevent online victimisation in the first place can only be effective if the consequences for its victims are understood and accepted. As such, it is vital to determine how online hate is perceived by online users before developing or adapting prevention strategies, such as educating users, accordingly.

5.2.4 The Role of Online Anonymity in Expressing Online Hatred

The role online anonymity plays in facilitating negative online behaviour has already been demonstrated in Computer Mediated Communication (CMC) research (e.g. Coffey & Woolworth, 2004; Doughty, Lawson, Linehan, Rowland & Bennett, 2014; Hardaker, 2010; Herring, Job-Sluder, Scheckler & Barab, 2002; Joinson, 1998; Lange, 2007; Moor et al., 2010; O’Sullivan & Flangin, 2003; Postmes, Spears, Sakhel & de Groot, 2001; Roberts & Wasieleski, 2012, Sia, Tan, & Wei, 2002; Steinfeldt et al., 2010; Suler, 2004; Turnage 2007). Negative online behaviours can include hateful expressions, such as flaming (outlined above) or trolling, which is defined as “*deliberately posting provocative and insulting messages about sensitive subjects or inflicting racism or misogyny on an individual*” (Internetmatters.org, n.d.). Negative online behaviour has been frequently attributed to online anonymity, which, not only removes social constraints (i.e. reduces inhibition; Fujita, Henderson, Eng, Trope & Liberman, 2006; Joinson, 1999; Suler, 2004), but it also provides users with the confidence to get away with such behaviour (Comas-Forgas & Sureda-Negre, 2010; Sawyer, 2016).

Yet, it is unclear whether people’s perceptions of their anonymity online have changed as a result of recent increases in media reports about online hate offences and the

introduction of modern technology (e.g. smartphones). In particular, Rohlffing and Sonnenberg (2016) observed that YouTubers removed hateful online content they had posted following potential legal threats by other YouTubers. Yet, online anonymity was not the specific focus of their study and thus was not investigated further. Nevertheless, the removal of hateful online content could be indicative of decreases in levels of perceived online anonymity, though, the rising number of online hate offences would suggest otherwise.

5.2.5 Research Questions.

Given the points above, this study is an exploration into SNS/SMP users' attitudes to the permissibility of online expressions of hate, providing an insight into the following research questions:

(1) Constitution of online hatred: Is there an overall consensus on what constitutes online hatred? Does the perceived hatefulness of online content correlate with its perceived criminality? Do perceptions of criminality correspond with legal definitions of online hate crimes? In line with this, it should be noted that some of the content described below contains offensive, racist and potentially upsetting.

(2) Online versus offline hatred: To what extent are online expressions of hate regarded to 'translate' into the offline world? Is online hate believed to be less bad/as bad/worse than offline hate? Should online hate be permitted as part of people's right to freedom of expression? To what extent are victims of online hate, Internet Service Providers (ISPs), the creator of hateful expressions, or the police thought to be responsible/accountable for and/or should deal with the consequences of online hate?

(3) Online anonymity: Has the introduction of modern technology changed the way SNS/SMP users feel about anonymity? Has an increase in news headlines stating the

consequences for online hate offences influenced how anonymous SNS/SMP users feel online? Is online anonymity perceived to affect users' online behaviour?

5.3 Method

5.3.1. Ethical Considerations

Participant consent was sought prior to participation, informing them of the nature and format of the study (i.e. a survey, exploring attitudes concerning online anonymity and whether perceptions of online hatred are in line with legislation). Participants were informed that they could withdraw from the study at any point up until its completion and provided them with the relevant contact details in case of any arising ethical concerns. They were also pre-warned of provocative statements prior to participation (see Appendix XIV), and were informed that participation was anonymous insofar that they did not have to reveal their names. After participation, participants received detailed debrief information about the study (see Appendix XVI).

5.3.2 Participants

Initially, 262 participants were recruited to complete an online survey measuring their attitudes towards the permissibility of online hatred. A total of 98 participants (i.e. 37%) were excluded because they did not complete all of the survey, leaving the data of 164 participants (79 females, 82 males, one transgender, one genderless and one omission) aged between 17 and 80 ($M = 33$, $SD = 16.44$) for the analyses. Eighty-five percent of participants identified as White, 4% as Mixed Race, 4% omitted indicating their ethnicity, 2% identified as Black, 2% as Asian, 2% as Middle Eastern, 1% as Indian. In addition, 97% of participants resided in the UK, 2% were non-UK based, and 1% did not indicate their place of residence. Finally, 43% were students, 41% were employed, 6% were unemployed, 6% were retired, 2% were self-employed and 2% omitted their work status.

5.3.2.1 Social networking. Sixty percent of participants indicated that they used their smartphones to connect to SNSs/SMPs. For a summary of participants' SNS/SMP activity see Table 5.1.

Table 5.1

Summary table illustrating key findings concerning participants' SNS/SMP activity.

Type of SNS	Access	Account activity	Activity
	Frequencies		Frequencies
Facebook	87%	Check account at least once a day	68%
Twitter	50%	Log into account at least once a day	23%
		Tweets per day	8%
LinkedIn	34%	Check account once a day	2%
		Check account monthly	16%
YouTube	76%	Use at least once a day	38%

5.3.3 Materials and Procedure

Participants were recruited online from the university student and staff population, as well as the general public (through personal contacts) and completed an online survey (via Qualtrics) in one single session. There was no time restriction and participants were able to omit responses. The survey first aimed to gather an overview of participants' demographic information and their social networking habits. It then focussed on three proposed attitude dimensions, namely, (1) *hatefulness and criminality*; (2) *online versus offline hatred*; (3) *the role of anonymity in expressing online hatred*; which were addressed across six attitude measures (outlined below). All measures were developed specifically for the purpose of this study. It should be noted that online users can perceive online hatred to be directed at different social categories (e.g. personally, peers, others). In this study,

most of the measures assessing participants' perceptions towards online hatred did not address a particular social category (apart from the last, namely, the *Online Anonymity*, measure, which addressed the participant directly; see Appendix XV). Therefore, the majority of results should be interpreted to predominantly address 'others'.

5.3.3.1 Hatefulness and criminality. This measure aimed to assess (i) what type of content constitutes online hatred; (ii) what content should be classified as criminal; (iii) whether a statement which is perceived as hateful is indicative of its perceived criminality; (iv) whether perceptions of the criminality of content are in line with its actual legislative status. The measure consisted of 18 attitude statements reflecting hateful/criminal and non-hateful/non-criminal online content (see below).

The attitude statements for this measure were selected from actual online content. Specifically, it was chosen from Rohlffing and Sonnenberg's (2016) data corpus (i.e. YouTube comments made in response to a video clip depicting racist behaviour exerted by a woman on a public train). To ensure representativeness, statements were selected by applying several search terms (including *Nigger, British, superior, ape, kill, death, Muslims, scum, Pikeys, bitch*) across the entire data corpus. In addition, in conjunction with an academic hate crime expert, particular attention was paid to the potential criminal implications of some of the statements during the selection process. This allowed the exploration of whether participants' perceptions of (il)legal web content matched current legislation, by comparing the actual legal status of statements with their perceived legal status. The selected statements were of similar word length (i.e. equally comprehensive).

The 18 statements were clustered as follows: (1) criminal - three statements met the legal threshold as inciting racial hatred (e.g. *Don't speak but kill all the non-Whites on tram fucking invaders*, Cronbach coefficient .89); (2) borderline criminal - four statements almost classified as meeting the legal threshold for the incitement to racial hatred (e.g. *So*

what if whites are descended from Africans? We're smarter, less prone to violence, and in every way superior. Black people are lesser evolved apemen, Cronbach coefficient, .90.); (3) offensive - nine statements were offensive but not legally defined as constituting a crime (e.g. *I'm okay with everyone.... except Muslims.. they do not know how to integrate and they always make a mess anywhere they go. and their fucking religion of Allahdin*, Cronbach coefficient, .90); (4) neutral - two statements were considered neutral (e.g. *Why can't we just stick together?! When will you people realise it is a colour and colour is not synonymous with being British.*). For a full list of statements from this attitude measure as well as the other five discussed below, see Appendix XV.

For each of the 18 statements, participants indicated the extent to which they considered the statement to be hateful, as well as the degree to which they considered the same statement as criminal (i.e. one response for each statement's level of hatefulness and one for its level of perceived criminality). Responses were given on seven-point Likert scales, ranging from 1-strongly disagree to 7- strongly agree with a neutral point (i.e. neither agree or disagree) of 4 in the middle. For example, two strongly disagree responses to the statement '*Trouble is... Blacks and Muslims in the 3rd world are lazy*', indicated that the statement was perceived to be not at all hateful and that it should definitely not be regarded as criminal. In addition, two strongly agree responses indicated that the statement was perceived to be very hateful and that it should definitely be regarded as criminal).

5.3.3.2 Effects of online hate on the offline world. This measure aimed to explore the overall relationship between online and offline hatred and specifically whether online hatred leads to offline hatred. This measure consisted of six attitude statements concerning online hatred in offline contexts (e.g. *Online posts, regardless of their content, are no more than words and can't hurt anyone in the real world; Social divisions are not related to anything said online*). Participants indicated their (dis)agreement to each statement on

seven-point Likert scales, ranging from 1- strongly disagree to 7- strongly agree with a neutral point (i.e. 4- neither agree or disagree) in the middle. Low scores here indicated that participants did not link online to offline hatred and high scores indicated that participants linked online to offline hatred. The remaining measures tapping the overall relationship between online and offline hatred used similar seven-point response scales.

Principal Component Analysis (PCA) revealed that five of the six statements showed an overall clear factor structure (i.e. factor loadings above .50 onto one factor, '*Online hatred does not transpire into offline hatred*'), accounting for 52% of the total variance). One statement (i.e. *Racist online posts can lead to social divisions in real life*) loaded negatively onto the factor and was reverse scored. The resulting measure showed good reliability (Cronbach's coefficient, .81).

5.3.3.3 Severity of online versus offline hate. This measure further aimed to assess the overall relationship between online and offline hatred and specifically explored whether the consequences of either online or offline hatred are perceived to be worse for its victims. The measure consisted of nine statements, which reflected the consequences of online and offline victimisation (e.g. *Offline hate is worse than online hate because it can involve physical confrontation; online hatred is worse than offline hatred because the revictimisation is ongoing*).

PCA revealed that eight of the nine statements showed a clear factor structure across two factors (i.e. factor loadings above .50; accounting for 54% of the total variance). Specifically, four statements loaded onto the first factor 'Offline hatred is worse than online hatred' accounting for 31% of the total variance with good reliability (i.e. Cronbach coefficient, .83). In addition, four statements loaded onto the second factor 'Online hatred is worse than offline hatred' accounting for 23% of the total variance with

an acceptable internal consistency (Cronbach coefficient, .68). The ninth statement was subsequently excluded from further analyses.

As a consequence, low scores on the seven-point response scales relating to statements concerning offline hatred to be worse than online hatred, indicated participants' agreement that the consequences of offline hatred are worse than those of online hatred and vice versa. In addition, low scores on the seven-point response scales relating to statements concerning online hatred to be worse than offline hatred, indicated participants' agreement that the consequences of online hatred are worse than those of offline hatred and vice versa.

5.3.3.4 Freedom of expression. This measure tapped the relationship between online and offline hatred by assessing whether expressions of online hatred would be acceptable as part of one's freedom of expression. The measure comprised seven attitude statements which addressed participants' stance on legislation governing expressions of online hatred (e.g. *Legislation should not dictate what people can or cannot say online*, or *People posting offensive things about others should be arrested*). Low scores to the seven-point response scales (i.e. here, strong agreement to the measures' statements) indicated that online hatred should be restricted by legislation and vice versa.

PCA showed a clear one-factor structure with factor loadings above .50 accounting for 46% of the total variance. Two statements (*People should be allowed to say whatever they want regardless of whether this is online or offline* and *Legislation should not dictate what people can or cannot say*) loaded negatively onto this factor and were subsequently reverse scored, resulting in good reliability (Cronbach coefficient, .80).

5.3.3.5 Responsibility-accountability. This measure also tapped the relationship between online and offline hatred. It specifically aimed to assess who participants hold responsible for the incitement of online hatred and who should deal with it. Specifically,

four statements attributed responsibility for online abuse to the victim (e.g. *It is your own responsibility to stop being bullied online by simply turning off the machine*; Cronbach coefficient, .79) and three to Internet Service Providers (*It is the Internet Service Providers' responsibility to remove hateful content that has been posted on their websites*; Cronbach coefficient, .84). In addition, one statement held the poster/creator of hateful content accountable (i.e. *It is not the poster's responsibility to consider how others might perceive their posts*) and one the police (i.e. *It's not up to the police to investigate people posting hateful things about others online*). Low scores to the seven-point response scales (i.e. here, strong agreement to the statements) placed the blame/responsibility for online hatred on the relevant party, whereas high scores removed the blame/responsibility from them. Please note that the wording of the statements blaming the police and the poster made it necessary to reverse the scores to keep in line with the remaining response scales.

5.3.3.6 Online anonymity. This measure aimed to assess the role of online anonymity in expressing online hatred by exploring (i) if participants feel anonymous online; (ii) whether perceptions of online anonymity have decreased over time; (iii) whether negative online behaviour is determined by perceived anonymity. The measure consisted of 26 attitude statements (e.g. *On the Internet, I feel completely anonymous*; *I'd remove some things I previously posted online if I thought they could be traced back to me*; *I felt more anonymous online before Facebook*.) Participants indicated their (dis)agreement to each statement on seven-point Likert scales, ranging from 1- strongly disagree to 7- strongly agree with a neutral point (i.e. 4 - neither agree or disagree) in the middle. Low scores (i.e. here, agreement with the statements) indicated (i) low levels of perceived online anonymity; (ii) reduced online anonymity over time due to technical advancements and recent media reports of online hatred (iii) anonymity determining (negative) online behaviour.

PCA revealed that twelve statements highly loaded onto one factor (i.e. above .50) accounting for 28% of the total variance. These statements referred to anonymity/identifiability affecting online behaviour and perceptions of reduced anonymity. One statement (i.e. *Anonymity doesn't determine what I say towards others online*) loaded negatively onto this factor and was subsequently reverse scored. This measure showed good reliability (i.e. Cronbach coefficient of .85). In addition, twelve of the 14 statements which did not load onto this factor were subsequently removed from further analyses. However, the two statements concerning participants' perceptions of online anonymity (i.e. one of the aims to be assessed with this measure) had to be retained.

5.3.4 Data Preparation and Transformation

First, the data were checked for missing and anomalous values, revealing that 13 participants had omitted responding to the online anonymity measure. These 13 were excluded from the analyses of this particular measure. In addition, inspection revealed that the highest percentage of missing values per statement across all attitude measures was 8%, which is lower than the common missing data rate of 15-20% observed in psychological studies (Enders, 2003). It is also lower than the rate of missing values which is likely to result in a biased statistical analysis (i.e. more than 10%, Bennett, 2001). Missing values were replaced via mean substitution. Next, participants' mean scores for all of the proposed measures were calculated, and these mean scores could range from 1 to 7. For the *hatefulness and criminality measure*, this meant that the mean scores for each level of legal status (i.e. offensive, borderline, criminal) was calculated. For the *severity of online hatred measure*, this involved the calculation of one mean score for each factor (i.e. offline hatred is worse than online hatred and online hatred is worse than offline hatred). For the *responsibility-accountability measure*, a mean score per relevant party (i.e. one for the victim, one for the poster, one for Internet Service Providers one for the police) were

calculated. For the online anonymity measure, a mean score for each area of exploration (i.e. perceptions of online anonymity/identifiability, whether online anonymity has decreased over time, or, whether online anonymity determines negative behaviour) was calculated. For the other measures (i.e. *effects of online hatred* and *freedom of expression*) this meant calculating one overall mean score per measure.

5.4 Results

The Results section is divided by research aims, namely ‘what content constitutes online hatred’, ‘online versus offline hatred’ and ‘online anonymity’ (see section 5.2.5), and so each will be examined in turn. Please note that all statistical analyses, means and frequencies in this Results section are based on participants’ overall mean scores in relation to the relevant variables/measures (see section 5.3.3). Means and frequencies will be reported in addition to any inferential statistics to illustrate participants’ overall perceptions of the different aspects of online hatred investigated in this study.

Frequencies in this Results section will be illustrated as ‘agreement’, ‘disagreement’ or ‘indifference’ to variables’/measures’ content. For the individual statements (e.g. responsibility of the poster or the police towards online hatred), the score classification was calculated so that scores between ‘1’ and ‘3’ represented ‘agreement’ (i.e. here, attributed responsibility to the relevant party), scores of ‘4’ ‘undecidedness’, and scores between ‘5’ and ‘7’ ‘disagreement’ (here, removed responsibility of the relevant party). To determine these categories across the measures (i.e. clusters of statements) scores were split and added together so that scores ranging from ‘1’ to ‘3.49’ represented ‘agreement’, ‘3.50’ to ‘4.49’ represented ‘indifference’, and scores between ‘4.50’ and ‘7’ represented ‘disagreement’. Moreover, differences within measures/variables (i.e. from neutral) will be determined with one-sample t-tests (i.e. by comparing the measures’/variables’ overall mean scores against their neutral scores).

5.4.1 What Content Constitutes Online Hatred?

5.4.1.1 Hatefulness. Frequency analysis of participants' agreement to statements' hatefulness ratings (i.e. here, scores between '4.50' and '7') indicated that the large majority of participants (i.e. more than 80%) rated all of the statements (offensive, borderline criminal, criminal) to be hateful. In addition, a one-way within subjects ANOVA was conducted to compare the differences in participants' mean perceived hatefulness ratings across three legal statuses (offensive, borderline criminal, criminal). There was a significant difference between participants' perceived hatefulness ratings and the statements' legal status, Wilks' $\lambda = .608$, $F(2, 162) = 52.26$, $p < .001$, $\eta_p^2 = .392$. Follow-up paired t-tests revealed that participants rated the offensive statements significantly less hateful compared to the borderline criminal ($t = 5.291$, $df = 163$, $p < .001$, CI [.16, .36], $d = .21$) or the criminal statements ($t = 10.064$, $df = 163$, $p < .001$, CI [.52, .78], $d = .58$). In addition, participants rated the borderline criminal statements as significantly less hateful than the criminal statements ($t = 5.883$, $df = 163$, $p < .001$, CI [.25, .52], $d = .33$). A summary of the mean scores and frequencies (including those of the criminality section, discussed next) can be found in Table 5.2 below.

Three one-sample t-tests (offensive, borderline criminal, criminal) showed that all three mean scores differed significantly from the neutral score of '4' (**offensive**: $t = 20.937$, $df = 163$, $p < .001$ CI [1.72, 2.08]; **borderline criminal**: $t = 21.846$, $df = 163$, $p < .001$ CI [1.97, 2.36]; **criminal**: $t = 29.823$, $df = 163$, $p < .001$ CI [2.38, 2.72]). These findings therefore suggest that although participants rated all statements, regardless of their legal status, to be hateful (i.e. above the neutral score of '4'), they nevertheless associated the statements in the rank order of their levels of criminality with current legislation.

5.4.1.2 Criminality. Frequency analysis of participants' agreement to statements' criminality ratings¹ (i.e. here, scores between '4.50' and '7') indicated that participants' perceptions of hateful content could match current legislation. In particular, although 38% of participants rated the offensive but legal statements to be criminal (i.e. not in line with legislation), 51% of participants rated the borderline criminal statements to be criminal and 83% of participants rated the criminal statements to be criminal. To test this further, a one-way within subjects ANOVA was conducted to compare the differences in participants' mean perceived criminality ratings across three legal statuses (i.e. offensive, borderline criminal, criminal). There was a significant difference between participants' perceived criminality ratings of the statements and their legal status, Wilks' $\lambda = .412$, $F(2, 162) = 115.45$, $p < .001$, $\eta_p^2 = .588$. Follow-up paired t-tests revealed that participants rated the offensive statements as significantly less severe compared to the borderline criminal ($t = 6.922$, $df = 163$, $p < .001$, $CI [.28, .50]$, $d = .22$) or the criminal statements ($t = 15.106$, $df = 163$, $p < .001$, $CI [1.34, 1.74]$, $d = .96$). In addition, it showed that participants rated the borderline criminal statements as significantly less severe than the criminal statements ($t = 11.688$, $df = 163$, $p < .001$, $CI [.96, 1.35]$, $d = .68$). These findings therefore illustrate that participants' perceptions of what content should or should not be classified as criminal do indeed mirror current legislation.

¹ Please note that the term criminality here encapsulates two concepts. First participants' perceptions of whether the content in question should be considered criminal or not to allow for the comparison between participants' perceptions of legality and content's actual legal status. Second, it referred to content's severity, as content can only be criminal or not. This is relevant to participants' criminality ratings across the three types of statements (offensive, borderline criminal, criminal).

Table 5.2

Summary table illustrating means, standard deviations and frequencies of participants' hatefulness and criminality ratings in relation to their legal statuses.

	Offensive (legal)			Borderline Criminal			Criminal		
	<i>M</i>	<i>SD</i>	Frequency	<i>M</i>	<i>SD</i>	Frequency	<i>M</i>	<i>SD</i>	Frequency
Hateful	5.90	1.16	83%	6.16	1.27	88%	6.55	1.10	95%
Criminal	4.25	1.64	38%	4.63	1.82	51%	5.78	1.56	83%

**Frequencies refer to participants' agreement that the statement content is hateful and should be criminal (here, scores above '4.49').*

5.4.1.3 The relationship between hatefulness and criminality. Correlations indicated that attitudes towards online hatred are a weak indication of their perceived criminality. In particular, there were weak positive correlations between participants' attitudes towards whether content was hateful and whether such content should be regarded as criminal for i) the offensive statements ($r = .419, N = 163, p < .001$), ii) the borderline criminal statements ($r = .388, N = 162, p < .001$), and iii) the legally criminal statements ($r = .348, N = 161, p < .001$). As such, these results suggest that although online content might be considered hateful and criminal, it tends to be considered more hateful than criminal.

5.4.2 Online versus Offline Hatred

5.4.2.1 Effects of online hate. Low mean scores and frequency analysis suggested that overall participants did not link online with offline hatred ($M = 2.25, SD = .97$, 89% of participants disagreed and only 2% agreed that online hatred has consequences for its recipients offline). A one-sample t-test comparing this mean score ($M = 2.25$) with the

neutral score '4' confirmed that participants perceived online hatred to have no effects on anyone in the offline world ($t = 22.99$, $df = 163$, $p < .001$, $CI [1.90, 1.60]$). These findings indicate that online users may view online hatred as acceptable, given that its consequences are perceived to only affect people online. For a summary of the findings concerning all perceptions of the relationship between online and offline hatred, see Table 5.3 below).

5.4.2.2. Severity of online versus offline hatred. Here, mean scores and frequency analysis around the neutral score of '4' showed that there was no overall agreement on whether the consequences of online or offline hatred are worse for its victims (offline worse: $M = 3.92$, $SD = 1.39$, 34% agreed and 35% disagreed that offline hatred is worse; online worse: $M = 3.96$, $SD = 1.08$, 30% agreed and 35% disagreed that online hatred is worse). Two one-sample t-tests (i.e. offline hatred is worse, online hatred is worse) showed non-significant differences between these scores and the neutral '4' score (**offline**: $t = -.701$, $df = 163$, $p = .484$, $CI [-.29, .14]$; **online**: $t = -.523$, $df = 163$, $p = .602$, $CI [-.21, .12]$), which confirms that overall participants neither agreed nor disagreed with the statements that offline hatred is worse than online hatred or that online hatred is worse than offline hatred. This was further supported when testing for a statistical difference between the two mean scores. That paired-sample t-test revealed a non-significant effect ($t = .233$, $df = 163$, $p = .816$, $CI [.24, .30]$).

5.4.2.3 Freedom of expression. High mean scores and frequency analysis revealed that most participants objected to legislation censoring online hate speech ($M = 4.64$, $SD = 1.10$, 60% of participants disagreed and 13% agreed that online hatred should be governed by legislation/ have legal consequences). A one-sample t-test confirmed that overall participants saw hate speech as part of freedom of expression (i.e. opposed its illegality; $t = 7.410$, $df = 163$, $p < .001$, $CI [.47, .81]$). These findings suggest that participants might

reject the concept of online victimisation constituting a criminal offence, which, in turn, further indicates that online hatred, directed at others, is perceived to be largely normative.

5.4.2.4 Responsibility/accountability. Mean scores and frequency analyses for the variables regarding to whom participants attributed the blame for and/or duty to intervene/remove hateful online content (thus ‘agreement’ here referred to this attribution) revealed that, by and large, participants blamed and attributed the responsibility to act upon online abuse to the victims (61%, $M = 3.05$ $SD = 1.31$). It also revealed that participants did not hold the poster responsible for how their posts are perceived by others (82%, $M = 5.71$, $SD = 1.52$). Moreover, it showed that most participants (57%) did not attribute the duty to remove and investigate hateful online content to Internet Service Providers (ISPs; $M = 4.5$, $SD = 1.5$) or the police (61% disagreed that it is up to the police to investigate people posting hateful things online; $M = 4.75$, $SD = 1.73$). Four one sample t-tests (i.e. victim, poster, ISPs, police) confirmed that all four mean scores differed significantly from the neutral score of ‘4’, thus confirming the findings of the descriptive analyses (**victim**: $t = 9.245$, $df = 163$, $p < .001$, $CI [.74, 1.15]$; **poster**: $t = 14.418$, $df = 163$, $p < .001$, $CI [1.48, 1.95]$; **ISPs**: $t = 4.262$, $df = 163$, $p < .001$, $CI [.27, .73]$; **police**: $t = 5.558$, $df = 163$, $p < .001$, $CI [.48, 1.02]$).

Table 5.3

Summary table of overall means, standard deviations and frequencies of participants' (dis)agreement to the proposed attitude measures.

Measure	<i>M</i>	<i>SD</i>	Frequencies		
			Agreement	Undecided	Disagreement
Effects of Online Hate (Link Between Online and Offline Hatred)	2.25	.97	2%	9%	89%
Severity of Online versus Offline Hatred (Offline Hatred is Worse Than Online Hatred)	3.92	1.39	34%	31%	35%
Severity of Online versus Offline Hatred (Online Hatred is worse than Offline Hatred)	3.96	1.08	30%	35%	35%
Freedom of Expression (Legislation Should Restrict Online Hatred)	4.64	1.10	13%	27%	60%
Responsibility-Accountability (Victim Blame)	3.05	1.31	61%	24%	15%
Responsibility-Accountability (Poster Blame)	5.71	1.52	10%	8%	82%
Responsibility-Accountability (ISPs Blame)	4.50	1.50	24%	19%	57%
Responsibility-Accountability (Police Blame)	4.75	1.73	26%	13%	61%

5.4.3 The Role of Online Anonymity in Expressing Hatred

5.4.3.1 Perceptions of online anonymity. Descriptive statistics (i.e. low mean scores and frequency analysis) revealed that most participants (86%) did not feel anonymous online ($M = 2.09$, $SD = .95$). These findings were supported by a one-sample t-test, which showed that participants' overall mean scores concerning their perceived online anonymity differed significantly from the neutral score of '4' ($t = 24.641$, $df = 150$, $p < .001$, $CI [1.75, 2.06]$). For a summary of all of the results referring to the role of anonymity in expressing hatred see Table 5.4 below.

5.4.3.2 Reductions in online anonymity over time. Here, descriptive statistics (means and frequencies) indicated that there was no clear direction (29% agreed, 40% were indifferent, 31% disagreed) concerning changes in perceived online anonymity over time (i.e. since the introduction of smartphones and/or increased publications of legal consequences for online hate offences; $M = 3.97$, $SD = 1.18$). A one sample t-test also confirmed that participants' overall mean scores relating to reduced anonymity due to modern technology and/or news headlines, did not differ from the neutral score of '4' ($t = .303$, $df = 150$, $p = .763$, $CI [.16, .21]$).

5.4.3.3 Online anonymity and subsequent behaviour. Descriptive statistics (i.e. low mean scores and frequencies) revealed that the majority of participants (57%) indicated that they would feel more inhibited online if they were identifiable ($M = 3.35$, $SD = 1.00$). These findings were supported by a one-sample t-test, which indicated that participants' mean scores, in relation to online anonymity determining their subsequent behaviour, differed significantly from the neutral score of '4' ($t = 7.893$, $df = 150$, $p < .001$, $CI [.48; .81]$).

Table 5.4

Summary table showing overall mean scores, standard deviations, frequencies, t- and p-values of participants' perceptions concerning three aspects of online anonymity.

Area of exploration	Frequencies						
	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	Agreement	Indifference	Disagreement
Perceptions of online anonymity	2.09	.95	24.641	< .001	3%	11%	86%
Reduced anonymity over time	3.97	1.18	.303	.763	29%	40%	31%
Anonymity determining subsequent behaviour	3.35	1.00	7.893	< .001	57%	30%	13%

5.5 Discussion

The aim of this study was to examine people's attitudes towards the permissibility of online hatred. In particular, the three-overall research aims of this study were to explore (i) what constitutes hateful online content; (ii) the relationship between online and offline hatred, and; (iii) the role of online anonymity in expressing online hatred, each of which will be discussed in turn.

5.5.1 What Online Content Constitutes Hatred?

The first research aim explored if there is an overall consensus on what online content is perceived to be hateful. It also tested the relationship between whether perceived hatefulness is indicative of perceived criminality (i.e. whether hatefulness correlates with criminality). Finally, it determined whether participants' perceptions of criminal web content correspond or not with existing online hate legislation.

Results revealed an overall agreement on which online content should, or should not be classified as hateful and/or criminal. In addition, although the legal but offensive, borderline criminal and criminal statements were all rated to be hateful, the level of perceived hatefulness corresponded with the statements' actual legal grading (i.e. offensive statements were rated the least hateful and criminal statements the most hateful). Similar results were obtained from the criminality ratings whereby offensive statements were rated to be the least severe and the illegal statements the most severe. Moreover, although participants perceived statements to be hateful and/or severely criminal the association between perceived hatefulness and perceived criminal severity was weak. In other words, although participants might consider content to be hateful, they do not necessarily consider it as criminal.

Taken together, these results imply that current legislation reflects which online content users perceive to be hateful and criminal. As such, these findings also suggest that online hatred is not seen to be normative or acceptable.

5.5.2 Online versus Offline Hatred

The second research aim was to explore the potential link between online and offline hatred. The measures used aimed to tap specifically (i) whether online hatred is perceived to be linked to offline hatred; (ii) whether the consequences of online or offline hatred are worse for its victims; (iii) whether online hatred is part of one's freedom of expression; (iv) who is responsible for the prevalence of online hatred.

Results showed that the majority of participants do not associate online with offline hatred, thus suggesting that participants distinguish between these two types of offending. Yet, participants' attitudes towards the severity of online versus offline abuse revealed that there is no real consensus on whether the consequences of either online or offline hatred are perceived to be worse. One reason for this ambiguity might stem from the specific online and offline hate-consequence examples used in this study. In particular, before completing the survey, participants might only have been aware of the obvious consequences of offline hate (e.g. physical confrontation), whereas some of the statements in the survey specifically pointed out the consequences of online hatred (e.g. ongoing victimisation, not knowing who the perpetrator is). Consequently, pointing out the negative consequences of both online and offline hatred may have led participants to rate the consequences of online and offline hatred as almost equally bad.

In addition, results showed that most participants objected to legal consequences for online hate speech. These findings suggest that online hate speech is indeed perceived to be permissible because it is seen to be part of one's right to freedom of expression (Moor et al., 2010).

Moreover, and interestingly, participants largely blamed the victims of online hatred for their abuse, thus mirroring existing research into victim-blame of other crimes (e.g. rape; Bieneck & Krahe, 2011; Suarez & Gadalla, 2010). These findings also suggest that participants might not have identified the statement-content as being aimed at themselves or their peers. In particular, previous research has identified that people are usually less empathetic towards out-group members (i.e. here, victims of online hatred), compared to in-group members (i.e. peers; Tarrant, Dazeley & Cottom, 2009).

Along similar lines and in line with previous findings (e.g. Avon and Somerset constabulary, Churchill, 2016), most participants did not hold the police or Internet Service Providers (ISPs) responsible for dealing with online hatred. Perhaps participants view online hatred to be acceptable and therefore consider that neither the police, nor ISPs should have to be responsible for dealing with it. Moreover, participants' overall indication that online hatred does not transpire into the 'offline' world could stem from a potential lack of experience with online victimisation (i.e. personal or peers).

Overall, these results suggest that online hatred is perceived to be somewhat acceptable and/or normative. In particular, and although there was no consensus on whether online or offline hatred is worse for its victims, the finding that most participants (i) did not link online to offline hatred; (ii) objected to legal consequences for inciting hateful web content; and (iii) blamed the victim but not the poster of hateful online content for online victimisation, all imply its perceived acceptability.

5.5.3 The Role of Online Anonymity in Expressing Hatred

The third research aim was to examine participants' attitudes towards online anonymity. In particular, it explored whether or not being anonymous affects online behaviour and whether the introduction of modern technology and publicised

consequences of online hate offences have changed participants' perceptions of their own level of online anonymity.

The results illustrate that few Social Networking Site (SNS) and Social Media Platform (SMP) users feel anonymous online. They also highlight that anonymity aids negative online behaviour and vice versa and thus support the large body of existing online anonymity research (e.g. Coffey & Woolworth, 2004; Doughty, Lawson, Linehan, Rowland & Bennett, 2014; Hardaker, 2010; Herring, Job-Sluder, Scheckler & Barab, 2002; Joinson, 1998; Lange, 2007; Moor et al., 2010; O'Sullivan & Flangin, 2003; Postmes, Spears, Sakhel & de Groot, 2001; Roberts & Wasieleski, 2012; Steinfeldt et al., 2010; Suler, 2004; Turnage 2007). They also reflect Comas-Forgas and Sureda-Negre's (2010) and Sawyer's (2016) findings that online anonymity grants those who express unacceptable (i.e. socially undesirable) views online the confidence to get away with it.

In addition, there was no obvious trend concerning changes in participants' levels of online anonymity over time. In particular, most participants were indifferent to whether modern technology, increases in online hate news headlines, or pressure on Internet Service Providers to reveal the identities of online hate offenders had changed their perceptions towards being anonymous online. One reason for these findings relates to the possibility that participants may never have felt anonymous online. Whilst the researcher was predominantly interested in changes in levels of online anonymity, this study did not specifically ask participants about their original levels of perceived online anonymity (i.e. when they started using the Internet).

Taken together, these results suggest that although most participants do not feel anonymous online, anonymity still plays a role in the way they behave online. Results also highlight that modern technology and increased publicity of the consequences of online hatred does not change how anonymous people feel online.

Overall, the results concerning the permissibility of online hatred examined in this study are somewhat inconclusive. In particular, whilst participants' ratings concerning the hatefulness and criminal gravity of specific content mirrored current legislation, the relationship between hatefulness and the criminal severity of such content was found to be weak. As such, whilst participants seem to agree on which content classifies as hateful and illegal, they still seem to distinguish between the hatefulness of such content and its criminal gravity, thus indicating a degree of permissibility towards such content. Another indicator towards the acceptability of hateful online content is reflected in participants' objection to legal consequences for the incitement of online hatred, and them blaming the victims of online hatred for their abuse. In particular, and when considering the finding that participants did not link online with offline hatred it could be that participants have come to accept/expect to experience some hateful online content which they then view as harmless and inconsequential. After all, this study did not obtain data on participants' history on online abuse and the pre-participation warning of the aims and the type of content would have made it highly likely that no victims of online abuse would have participated in this study. Another indication against the permissibility of online hatred and negative online behaviour relates to participants' acknowledgement that they would only be inclined to behave 'badly' online if they were anonymous. Yet, it needs to be noted that although standing against the permissibility of online hatred in general, anonymity has been linked with other underlying concepts (e.g. no social constraints/reduced inhibition; Fujita, Henderson, Eng, Trope & Liberman, 2006; Joinson, 1999; Suler, 2004) and confidence to get away bad behaviour; Comas-Forgas & Sureda-Negre, 2010; Sawer, 2016). As such, there is no straight forward link between anonymity and the permissibility of online hatred.

5.5.4 Limitations

As with survey data in general, one of the limitations of this study concerns its high participant exclusion rate (i.e. 37%). In particular, this study's exclusion rate was based on participants not completing the entire survey. Whilst one can anticipate that not every participant will complete the entire survey, the sensitive context of this survey (i.e. socially undesirable behaviour online) may have further contributed towards a relatively high non-completion rate. Specifically, and despite prior warning, perhaps participants became offended by the hateful nature of the statements employed as measures here, or by disclosing information about their online behaviour. Evidence for this possibility comes from the large number of participants who did not respond to any of the questions (i.e. 28%). As such, one could argue that the participant sample is not representative and thus the findings of this study are not readily generalisable.

In addition, although this study used real-life examples of hateful online statements to explore whether there is an overall consensus on what constitutes online hatred, it still lacks ecological validity. Specifically, the survey used self-reports to assess participants' attitudes towards a range of online hate-related variables. There are two issues with this method. First, the method of self-reporting has been criticised for its subjective nature (Greenwald & Banaji, 1995; Gawronski, Galdi & Arcuri, 2015). Second, the hypothetical nature of the statements to which participants responded does not allow us to draw any conclusions about 'actual' online hate offenders' attitudes concerning online hatred.

Furthermore, as previously mentioned, most of the statements did not specifically address participants themselves, or their peers. Instead, they addressed non-specified others, which not only prevents conclusions about who exactly participants had in mind when rating the statements, but may also have resulted in participants showing less empathy towards victims of online hatred (e.g. Tarrant et al., 2009). For ethical and legal

reasons this study did could not ask participants if they had either been victimised, witnessed or exerted hate online. As such, and as pointed out before, the sample cannot be considered as representative.

Moreover, and as indicated above, the term permissibility used throughout this Chapter is somewhat problematic as some of the aspects explored seem to relate closer to the concept of acceptability of online hatred rather its permissibility. As such, some of the results (e.g. participants' opposition to legislating hateful online content, and their attribution of blame/responsibility/accountability) need to be interpreted with some caution.

Finally, this study did not obtain data on participants' actual (posting) activity/frequency on SNSs/SMPs. Instead, the most frequent response option referred to *several times a day*. The interpretation of participants' account activity (see Section 5.3.2.1) is therefore somewhat problematic as it might not reflect their actual SNS/SMP activity. In particular, users might post things on SNSs/SMPs every hour or even more often, whereas others 'only' two times a day. As such, combining the posting frequencies would have automatically resulted in losing some of the depth and range of the data.

5.5.6 Future Directions

5.5.6.1 Changing attitudes. In line with online victimisation, future research should explore how to change online users' attitudes towards the consequences of online victimisation. In particular, this study illustrates that although there was an overall consensus on what constitutes online hatred, the results of several measures indicated that participants perceived online hatred to be largely acceptable. It would be interesting to examine if hateful content which addresses participants personally or their peers would elicit similar levels of perceived permissibility of online hatred. Such research could also determine whether current perceptions of the consequences of online hatred (i.e. perceived

harmlessness and inconsequentiality) are based on unawareness/a lack of understanding of the magnitude for its victims, or on a lack of recognition/empathy.

Along similar lines, future research should compare the attitudes of online hate offenders, victims of online hatred and ‘everyday’ online users towards the consequences of online hate offences. Such research would not only aid the understanding of the similarities and differences in how online hate is expressed and perceived by online users, but it could also be used to develop strategies to raise awareness of the consequences online hatred has for its victims. Therefore, when raising awareness of the consequences of online hatred, existing attitudes towards the effects of online hatred might also change and thus prevent people from inciting online hatred in the future. This would be particularly relevant for those who blame the targets of online hate for their own victimisation. As such, changing attitudes might not only help the victims’ recovery, in terms of recognition and the reduction of possible feelings of self-blame, but it could also help some of those who incite online hatred in the process of their rehabilitation (e.g. realise the gravity of their offence).

However, such interventions also need to consider/alter the underlying motivations of hate offenders. Specifically, they need to consider if an offender offends to seek a thrill from it, to defend their territory (e.g. due to perceptions of threats to their resources), to retaliate in response to perceived degradation or assaults on their in-group, or because they want to rid the world of felt evil or inferior groups (McDevitt, Levin, Bennett, 2002). Such motivational distinctions are important, as offenders might be aware of the seriousness of their actions but might deem it necessary to achieve their goals. Moreover, the recognition of the magnitude of online hate offending might prevent those who have not yet incited online hatred from creating such content in the first place.

5.5.6.2 Online anonymity. Despite the findings that perceptions of online anonymity levels have decreased for almost half of the participants, the number of online hate offences are increasing. Future research should explore this association more closely. In particular, whilst this study was able to attribute decreases in perceived online anonymity to smartphones and media attention (i.e. fear of legal consequences), it is not possible to make inferences about the specific motivating factors in increases in online hate offending. Thus, future research should explore any additional motivating factors that can facilitate online hate offending.

5.5.7 Conclusion

We increasingly encounter hate on the Internet in our daily online business. In order to counter online hatred effectively, it is important that we know how it is perceived by ‘everyday’ users. The results of this study, concerning the permissibility and acceptability of online hatred, are somewhat inconclusive. Whilst participants did not connect online hatred to offline hatred, opposed its legislation, deemed it inconsequential to its victims and even held the victims responsible for it, hateful content is rated to be not only hateful but also criminal and participants admitted that they would only behave badly online if they were anonymous. Whilst the sample used in this study does not allow for the findings to be easily generalised they, nonetheless, suggest ideas for future research aimed at raising awareness of the consequences for victims of online hatred to try and change existing attitudes towards the incitement of online hatred.

5.6 References

- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology, 41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Association of Chief Police Officers (2014). *ACPO hate crime manual*. ACPO: London.
- BBC News (2016, January 27). Quarter of people have witnessed hate crime, poll suggests. *BBC News*. Retrieved from <http://www.bbc.com/news/uk-35415988>
- Bennett D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*, 464–469. DOI: 10.1111/j.1467-842X.2001.tb00294.x
- Berliner Zeitung (2016, October 16). Volksverhetzung- Berliner in zu Bewährungsstrafe wegen Hass-Kommentaren auf Facebook verurteilt. *Berliner Zeitung*. Retrieved from <http://www.berliner-zeitung.de/berlin/volksverhetzung-berlinerin-zu-bewaehrungsstrafe-wegen-hass-kommentaren-auf-facebook-verurteilt-23043144>
- Bieneck, S., & Krahe, B. (2011). Blaming the victim and exonerating the perpetrator in cases of rape and robbery: Is there a double standard? *Journal of Interpersonal Violence, 26*, 1785-1797. DOI: 10.1177/0886260510372945.
- Burlacu, A, (2016, August 27). Microsoft tackles online hate speech with new tools and resources to combat abuse. *Tech Times*. Retrieved from <http://www.techtimes.com/articles/175274/20160827/microsoft-tackles-online-hate-speech-with-new-tools-and-resources-to-combat-abuse.htm>

- CBC News (2015, November 12). Arthur Topham convicted of promoting hatred against Jews on website. *CBC News*. Retrieved from <http://www.cbc.ca/news/canada/british-columbia/topham-anti-semitism-jewish-hatred-1.3317052>
- Churchill, L. (2016, August 17). Police asked Twitter users about online hate crime and did not get the response they expected. *Bristol Post Online*. Retrieved from <http://bristolpost.co.uk/twitter-users-deem-online-posts-stirring-up-hatred-as-banter-in-police-survey/story-29628587-detail/story.html>
- Comas-Forgas, R., & Sureda-Negre, J. (2010). Academic plagiarism: Explanatory factors from students' perspective. *Journal of Academic Ethics*, 8, 217–232. DOI: 10.1007/s10805-010-9121-0
- Corfield, G. (2016, June 2). Arrests for 'offensive' Twitter and Facebook messages up by a third: Said something naughty in London? You could be cuffed. *The Register*. Retrieved from http://www.theregister.co.uk/2016/06/02/social_media_arrests_up_37pc_london_section_127_communications_act/
- Crown Prosecution Service (2016, August 5). Anjem Choudary and Mohammed Rahman charged for inviting support of ISIL. *CPS.ORG.UK*. Retrieved from http://www.cps.gov.uk/news/latest_news/anjem_choudary_and_mohammed_rahman_charged_for_inviting_support_of_isil/index.html
- Enders C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, 8, 322–337. DOI: 10.1037/1082-989X.8.3.322.

- European Commission (2015, December 3). EU Internet forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online. *New Europe*. Retrieved from <https://www.neweurope.eu/wires/eu-internet-forum-bringing-together-governments-europol-and-technology-companies-to-counter-terrorist-content-and-hate-speech-online/>
- European Commission (2016, September 26). United against hate speech on the Web: where do we stand? – Speech by Commissioner Jourová at conference with German justice minister Maas. *New Europe*. Retrieved from <https://www.neweurope.eu/press-release/united-against-hate-speech-on-the-web-where-do-we-stand-speech-by-commissioner-jourova-at-conference-with-german-justice-minister-maas/>
- Foxman, A. H., & Wolf, C. (2013). *Viral hate containing its spread on the internet*. New York: Palgrave Macmillan.
- Fujita, K., Henderson, M. D., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science*, *17*, 278–282. Retrieved from <http://psych.nyu.edu/tropelab/publications/Fujitaetal2006.pdf>
- Gawronski, B., Galdi, S., & Arcuri, L. (2015). What can political psychology learn from Implicit measures? Empirical evidence and new directions. *Political Psychology*, *36*(1), 1-17. DOI: 10.1111/pops.12094.
- Graham, L. (2016, February 9). 24% of young people targeted by online abuse in the last year. *CNBC*. Retrieved from <http://www.cnn.com/2016/02/09/24-of-young-people-targeted-by-online-abuse-in-the-last-year.html>

- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27. Retrieved from http://faculty.washington.edu/agg/pdf/Greenwald_Banaji_PsychRev_1995.OCR.pdf
- Home Office (2009). *The cross-government action plan*. London: Home Office.
- Internetmatters.org (n.d.). *Definition of trolling*. Retrieved from https://www.internetmatters.org/issues/cyberbullying/learn-about-it/?gclid=CjwKEAiA6YDBBRDwtpTQnYzx5IASJAC57ObMOiqAxrJvSDYxjtmFXnYguQDgiBkwPc0JsP0ltX_H_RoCN0nw_wcB#1473780018865-0c69a128-9468
- Jacks, W. & Adler, J. R. (2015). A typology of online hate crime. *Open Access Journal of Forensic Psychology*, 7, 64-89. Retrieved from <http://eprints.mdx.ac.uk/18930/1/Jacks%20and%20Adler%202015.pdf>
- Jewish News Online (2015, December 17). Extremist Bonehill jailed for three years for inciting racial hatred. *Jewish News Online*. Retrieved from <http://jewishnews.timesofisrael.com/extremist-bonehill-guilty-of-inciting-racial-hatred/>
- Jewish Telegraphic Agency (2016, October, 28). Kuwaiti student convicted of hate crime over anti-Semitic Facebook post. *Jewish Telegraphic Agency*. Retrieved from <http://www.jta.org/?s=jumaa&orderby=date&order=desc>
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31, 433-438. DOI: 10.3758/BF03200723
- Kazakhstan Newsline (2015, August, 3). New person sentenced for ‘inciting religious hatred’ in Kazakhstan. *Kazakhstan Newsline*. Retrieved from <http://www.newsline.kz/news.php?y=15&m=08&d=03#0>

- Laubner, A. (2016, September 27). Prävention beginnt im Kopf- Islamwissenschaftlicher erläutert Wurzeln und Ursachen von Selbstradikalisierung. *Mindener Tageblatt*, 226, 8.
- Mail Online (2016, March 18). Dutch lawmaker Wilders in court on hate speech charges. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/wires/ap/article-3498381/Dutch-lawmaker-Wilders-court-hate-speech-charges.html>
- Mail Online (2016, March 24). Singapore jails Australian for provoking anti-foreign hatred. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/wires/ap/article-3507850/Singapore-jails-Australian-provoking-anti-foreign-hatred.html>
- McDevitt, J., Levin, J., & Bennett, S. (2002). Hate crime offenders: an expanded typology. *Journal of Social Issues*, 58, 303-317. DOI: 10.1111/1540-4560.00262.
- North, A. (2016, September 26). What it's like to fight online hate. *The New York Times*. Retrieved from http://mobile.nytimes.com/blogs/takingnote/2016/09/26/what-its-like-to-fight-online-hate/?_r=0&referer=https://www.google.com/
- Organisation for Security and Co-operation in Europe (2010). *Incitement vs. Freedom of Expression: Challenges of combating hate crimes motivated by hate on the Internet*. Report of the OSCE-ODIHR Expert Meeting in Warsaw. Retrieved from <http://www.osce.org/odihr/68750>
- Oryszczuk, S. (2016, August 16). Community welcomes Met's £1.7m unit to tackle online hate. *Jewish News Online*. Retrieved from <http://jewishnews.timesofisrael.com/community-welcomes-mets-1-7m-unit-to-tackle-online-hate/>
- Phillips, T. (2015, December 22). Pu Zhiqiang given three-year suspended sentence. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2015/dec/22/pu-zhiqiang-chinese-human-rights-lawyer-sentenced-to-three-years>

- Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: the effects of anonymity on group behaviour. *Personality and Social Psychology Bulletin*, 27, 1243-1254. DOI: 10.1177/01461672012710001
- Pradeep (2016, August 26). Microsoft is launching a new dedicated web form for reporting hate speech. *MSPoweruser.com*. Retrieved from <https://mspoweruser.com/microsoft-is-launching-a-new-dedicated-web-form-for-reporting-hate-speech/>
- Radio Free Europe Radio Liberty (2015, July 31). Another Kazakh sentenced for inciting ethnic hatred online. *Radio Free Europe Radio Liberty*. Retrieved from <http://www.rferl.org/a/kazakhstan-ethnic-hatred-man-sentenced/27163239.html>
- Reuters (2015, December 30). Russian court jails blogger for five years for ‘extremist’ posts. *Reuters*. Retrieved from <http://uk.mobile.reuters.com/article/idUKKBN0UD16O20151230>
- Roberts, J. A., & Wasieleski, D. M. (2012). Moral reasoning in computer-based task environments: exploring the interplay between cognitive and technological factors on individuals’ propensity to break rules. *Journal of Business Ethics*, 110, 355-376. DOI:10.1007/s10551-011-1196-z
- Sawer, P. (2016, September 29). Cyberbullying victims speak out: ‘they were anonymous so they thought they could get away with it.’ *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/technology/facebook/8885876/Cyberbullying-victims-speak-out-they-were-anonymous-so-they-thought-they-could-get-away-with-it.html>
- Sia, C. L., Tan, B. C. Y., & Wei, K. K. (2002). Group polarization and computer mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, 13, 70–90. DOI: 10.1287/isre.13.1.10.97

- Spencer, R. (2015, December 31). Kazakhstan: Convert from Islam to Christianity gets 2 years prison for inciting religious hatred. *Jihadwatch.org*. Retrieved from <https://www.jihadwatch.org/2015/12/kazakhstan-convert-from-islam-to-christianity-gets-2-years-prison-for-inciting-religious-hatred>
- Suarez, E., & Gadalla, T. M. (2010). Stop blaming the victim: A meta-analysis on rape myths. *Journal of Interpersonal Violence, 25*, 2010-2035. DOI: 10.1177/0886260509354503
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*, 321-326. DOI: 10.1089/1094931041291295.
- Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for out-group members. *British Journal of Social Psychology, 48*, 427-446. DOI: 10.1348/014466608X373589
- Taylor, R. (2016, September 1). Suspected neo-Nazi is accused of terror offence for 'posting picture of Adolf Hitler with 'kill the Muslims' message online'. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/news/article-3769231/Suspected-neo-Nazi-accused-terror-offence-posting-picture-Adolf-Hitler-kill-Muslims-message-online.html>
- Tech City News (2016, October 4). New tool to detect and prevent hate speech online. *Tech City News*. Retrieved from http://techcitynews.com/press_release/new-tool-detect-prevent-hate-speech-online/
- Wolfram, J. (2016, July 17). Can Facebook flush out the media trolls in the war of hate? *Recode*. Retrieved from <http://www.recode.net/2016/7/11/12123318/facebook-media-trolls-hate-speech-online-ab>

Chapter 6

General Discussion

6.1 General Discussion Outline

This chapter first summarises the main findings of the thesis with regard to how Social Networking Sites (SNSs) and Social Media Platforms (SMPs) can facilitate the spread of online hatred, how online users respond to online hatred, how online hatred is viewed by ‘everyday’ online users and how these findings fit within existing research (see Section 6.2). Next, the learning gains of the research programme and researcher biases on the findings are reflected upon (see Section 6.3). Then the practice, policy and research implications of the findings are considered, taking a focus on how hatred might spread and conversely, how it might be prevented. Here, the particular foci are a) governmental and Internet Service Providers (ISPs) strategies to counter online hatred, b) the role of social identification and online anonymity in accepting and expressing hatred and c) raise awareness to aid the prevention of online hatred and rehabilitate offenders (see Section 6.4). Next, some methodological issues are considered (see Section 6.5). Finally, ideas for future research are discussed which draw together and build on those that are mentioned in previous chapters (see Section 6.6) before drawing an overall conclusion (see Section 6.7).

6.2 Overview of Main Findings

To date, online hate research has primarily focused on individuals and groups who have already expressed hate on the Internet. Hence there is limited empirical evidence investigating how online hatred might affect ‘everyday’ users who do not (perhaps yet) seem polarised in their views. Specifically, we know little about the extent to which hateful online content on SNSs may shape people’s attitudes towards others. Therefore, the main aim of this thesis was to examine *if and how online hatred can influence SNS users’ attitudes (e.g. prejudice, expressions of hatred) towards others*. The studies conducted here focused on different aspects of online hatred and explored: (a) the influence of online

discussions on participants' levels of prejudice; (b) SNS users' responses when exposed to hateful online content and (c) the permissibility of expressions of hatred online.

Chapter 2 described two studies in which an explicit 10-item measure was developed to assess prejudice towards Roma and Travellers (the Anti-Roma and Traveller - ART scale). In doing so, an existing explicit prejudice measure (i.e. Levinson & Sanford's 1944 Anti-Semitism Scale) was successfully adapted. The newly-developed prejudice measure was intended to provide a tool for tracking changes in participants' levels of prejudice towards Roma and Travellers (see Chapter 3).

Chapter 3 examined the influence of online group discussions (via instant messaging) on participants' levels of prejudice towards a particular target group, namely Roma and Travellers. The experimental study protocol involved the measurement of participants' pre-existing levels of prejudice towards Roma and Travellers via completion of the ART scale (developed in Chapter 2). Next, participants with similar levels of prejudice (e.g. low, intermediate, high), plus one confederate, were grouped together before discussing, via instant messaging (i.e. typing comments online simultaneously), the 'real-world' Dale Farm case. The case concerned the eviction of Roma and Travellers from a particular plot of land (Dale Farm). During these discussions, the confederate responded with (typed) pre-determined statements which, depending on the experimental condition, aimed to either reduce or increase participants' existing levels of prejudice towards Roma and Travellers. After the online discussion, participants completed the ART scale again to determine any changes in levels of prejudice (i.e. online influence).

Overall, most participants resisted these attempts of online influence. However, participants with intermediate levels of pre-existing prejudice towards Roma and Travellers indicated significantly higher levels of prejudice following the discussion during which the confederate tried to increase their levels of prejudice. Yet, this influence was

one-directional, as those participants with intermediate levels of pre-existing prejudice towards Roma and Travellers who encountered the confederate who expressed views aimed at decreasing their levels of prejudice, resisted online influence. These findings are supported by previous research suggesting that those with strong attitudes (i.e. those with low or high levels of pre-existing prejudice towards the target group) will resist influential attempts which oppose their existing beliefs, as opposed to those with 'weak' attitudes (i.e. here, those with intermediate levels of pre-existing prejudice towards Roma and Travellers; e.g. Ajzen & Fishbein, 2005; Fazio, 1990; O'Keefe, 2004; Petty & Krosnick, 2014). The findings are also in line with research that suggests that those who have strong attitudes and are exposed to attitudes which run in line with their existing position, have no reason to change their position (e.g. Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Petty & Cacioppo, 1981; Sunstein, 2001; Wojcieszak, 2010).

It was surprising to find that participants with intermediate initial levels of prejudice who were exposed to experimental messages aiming to reduce their existing levels of prejudice towards Roma and Travellers, would resist online influence. This finding seems to suggest that pre-existing negative stereotypes associated with Roma and Travellers may be easier to reinforce than to refute. Overall, this chapter provides evidence that online interactions may, indeed, shape negative attitudes in some individuals. Specifically, influence seems to be determined by initial attitude strength, the direction of social influence (i.e. towards an increase or decrease of prejudice) and, arguably, the strength and, therefore the resistance to change, of the original stereotype(s).

The qualitative study described in Chapter 4 explored how YouTubers responded (via online comments) to specific online hate materials. In particular, YouTubers commented on a particular YouTube video clip, depicting racial abuse exerted by a White female woman called Emma. The analysis of the comments involved an inductive, data

driven approach using Thematic Analysis (TA; Braun & Clarke, 2006), which identified four main themes.

The first theme, called *Making Sense of Emma*, captured YouTubers' attempts to find explanations for Emma's behaviour. This was repeatedly linked to her perceived role as an unfit mother and her bad mental state and thus not only attempted to 'make sense' of Emma's behaviour, but also corresponds to Lloyds theory of double deviance (Heidensohn, 1989; Lloyd, 1995). In particular, Heidensohn (1989) and Lloyd (1995) found that women who offend (here, racially abuse others) are not only punished for their offence but also for their deviation from their gender expectations and social norms.

The second theme was called *Meeting Hatred with Hatred*, and described YouTubers' efforts to oppose racism whereby they resorted to aggressive and/or hateful language and thereby fits into existing flaming research (e.g. Coffey & Woolworth, 2004; Doughty, Lawson, Linehan, Rowland, & Bennett, 2014, Lange, 2007; Moor, Heuvelman, & Verleur, 2010). Flaming involves people's use of aggressive, abusive, and racist language to oppose certain views (here, racism). Such motivation is also closely linked to one of the typologies of hate offenders, namely, the retaliatory one, as identified by McDevitt, Levin and Bennett (2002). It also relates to another of these typologies, namely, the thrill seeing one (McDevitt et al., 2002), whereby some YouTubers may have responded in a hateful manner because they were seeking a thrill from their own and other responses (i.e. similar to the concept of trolling).

The third theme, namely *Us versus Them*, encapsulated YouTubers' leanings to categorise themselves and other YouTubers into in- and outgroups, which depended on their particular stance on racism. Such categorical thinking (see also section 6.4.2 – *The Role of (Social) Identification in Shaping Hatred* below) mirrors existing research from the broader social identity perspective (e.g. Angie et al., 2010; Blackwood, Hopkins &

Reicher, 2013, 2015; Bliuc, McGarty, Hartley & Muntele Hendres, 2012; Reicher, Haslam & Rath, 2008). It also and links to so called dichotomous (Black and White) thinking, whereby others (here, other YouTubers) were either supporting or opposing racist views. Dichotomous thinking, by means of identifying and positioning certain others outside the law (here, expressed as in-group favouritism and out-group discrimination), has also been classified as the first stage of genocide (Arendt, 1951) and ‘the fourth floor of the staircase to terrorism’ (Moghaddam, 2005).

The last theme was called *Contesting Britishness*. Here, YouTubers attempted to define (and challenge) what it means to be British. The contestation of one’s national identity (Branscombe, Ellemers, Spears & Doosje, 1999; Condor, 2000; Faulkner & Bliuc, 2016; Reicher et al., 2008) also fits within existing research from the broader social identity perspective (see section 6.4.2 *The Role of (social) identification in shaping hatred* below). Moreover, for those supporting racist views, the notion of Britishness was frequently linked to perceptions of an unjust denial of ones’ Britishness by the government due to immigration, resulting in a sense of grievance. This sense of grievance also links to the so called ‘mission’ and ‘defensive’ types of hate crime offenders (McDevitt et al., 2002), whose motivations to offend is based on them wanting to protect their home (here, the UK and their Britishness) from outsiders, including wanting to eliminate all evil (here, immigrants and immigrant supporters).

Overall, the analysis therefore highlighted that, in part, hateful web content can indeed fuel aggressive and hateful responses and that motivations towards posting such responses link to those of hate offenders and other offending research (i.e. female offending). Yet, it also showed that the large majority of comments emphasised collective group memberships, as well as people’s right to claim membership in a particular social category (here, Britishness). These findings therefore suggest that responses to online

racial hatred might involve broader social psychological processes rather than those who are exposed to it automatically endorsing it. In other words, there is not always a straightforward or inevitable escalation of hatred.

Chapter 5 explored the permissibility of online hatred across three areas of interest: (1) what constitutes online hatred; (2) the relationship between online and offline hatred; and (3) the role of online anonymity in expressing online hatred. Results revealed that most participants did not connect online with offline hatred, blamed the victim of online abuse but not the poster of hateful web content, rejected police involvement in cases of online hatred, and opposed legal consequences for the incitement of hatred- all pointing towards the permissibility of online hatred. Yet, results also showed that participants overall agreed on what content is perceived to be hateful and what content should be classified as criminal, whilst also acknowledging that online anonymity aids their own anti-normative behaviour- suggesting online hatred is not permissible. Therefore, the overall evidence concerning the permissibility of expressions of online hatred was not entirely clear. Overall, this study emphasises the complexities associated with the permissibility of online hatred.

Taken together, the evidence provided by the studies in this thesis suggests that expressions of online hatred can influence the attitudes of some online users. In particular, the current evidence suggests that online group discussions can increase prejudice in people whose initial attitudes are 'weak'. However, exposure to hateful online content does not automatically result in further hatred - instead, responses to online hate content can focus upon broader social psychological issues (i.e. social identification, including defining and contesting one's national identity, as well as positioning oneself and others according to their perceived social categories). In terms of the permissibility of expressions of online

hatred, the evidence obtained in this thesis is mixed because some findings point towards its permissibility, whilst others imply the opposite.

The diversity of the studies in this programme of research (e.g. explicit prejudice measure development, experimental and ‘observational’ approaches, qualitative and quantitative analytic methods) tapped a variety of questions concerning online hatred. As such, these findings further our particular understanding of how online hatred on SNSs may be used by polarised groups to fuel hatred and to effectively influence online users’ views.

6.3 Reflexivity and Learning Gains

In the beginning of this research programme, the complexity online hatred presents was rather overwhelming. It was particularly challenging to find a way to limit, yet maximise, the scope of the research programme to a number of aspects exploring online hatred. It soon became clear that one PhD can assess only so much and that the topic of online hatred warrants enough scope for far more research studies than fit into one PhD. For example, and although the idea of the study discussed within Chapter 3 was developed at the very beginning of this research programme, the careful consideration for appropriate target groups resulting in the need for an additional tool to measure online influence (here, changes in levels of prejudice) highlights only one aspect of the complexities this research programme faced.

However, this complexity offered me the opportunity to apply and gain experience in a range of different research methods. It also allowed for the critical evaluation of the methodologies, which were applied throughout this programme of research. As a result, it is now apparent that the development process of the explicit Anti-Roma and Traveller (ART) prejudice scale based on Levinson and Sanford’s (1944) Anti-Semitism (AS) scale did not require to make it generic/group non-specific (Study 1a)

before adapting it to the intended target groups. This is especially relevant, given that the AS scale was selected, in part, for the said overlap in stereotypes between Jewish people and Roma and Travellers (Barnett, 2013).

In addition, the completion of the study described in Chapter 3 reinforced the importance of keeping study designs as simple as possible. In particular, the combination of a complex design (e.g. its completion in two sessions, the need for participants with similar levels of prejudice towards the target groups, the organisation of several participants and one confederate to complete the task remotely, yet at the same time) and the type of task (i.e. discussing an unknown topic with unfamiliar others) resulted in difficulties to recruit participants and thus a prolonged data collection process. These difficulties also resulted in a change of the original study design. In particular, the original design included an additional type of argument, termed ‘neutral’ (see Appendix X). This manipulation was subsequently dropped as, in addition to the difficulties in recruitment, no online influence was anticipated from it. Whilst this condition could have served as a control condition to the other two types of arguments (pro-, anti-target groups) its retention would probably have resulted in a still ongoing data collection process. Despite these difficulties, the unique and new set-up made it a very interesting and positive experience with some interesting observations (e.g. who was influenced/resisted online influence, the role of online anonymity in negative online behaviour and its relation to the permissibility/acceptability of online hatred). Some of these ideas were then incorporated in the study described in Chapter 5 and some ideas considered for future research (e.g. reduced anonymity).

Moreover, this research programme provided me with the chance to gain experience in qualitative research with large amounts of data (Chapter 4). The manual coding of the data was informative and helpful for my engagement with the vast range of

responses, the development of an initial coding scheme, and the appropriate analytic approach, yet, it was also very time consuming. Therefore, if I was to repeat the process or undertake a similar research project, I would consider using data mining software, such as NVIVO.

Part of the qualitative research experience also included the acknowledgement of my own preconceptions which influenced my methodological and analytic decisions. For example, as a White German female, coming from a working-class background, I have long been interested and aware of the consequences of hate. Because of my lack of first-hand experience with hatred (i.e. projected against me or others), I found some of the hateful materials and online content disturbing and startling, particularly at the beginning of this research programme. In order to get somewhat desensitised to the nastiness of some of the materials and maintain a positive psychological well-being, I maintained a degree of distance to the content, which I achieved mostly through my role as a researcher. However, certain content continued to be disturbing, which is when I had to remind myself of the somewhat 'bleak' nature of the research topic. I also adopted a legal stance to deal with particular nasty the content, whereby the categorisation into offensive and possibly illegal content helped to rationalise that, although content might be horrid, there might be no element of illegality per se. This process further helped me to keep any negative arising emotions at bay. As a result, and despite my attempts to keep an emotional balance between distancing myself from the nastiness of the content and trying to maintain my privately held views (e.g. an awareness of how such content might make those addressed by it feel and general wrongness), these preconceptions ultimately influenced how I approached the data.

In addition, this emotional balance permitted me to support my confederates during and after their attempts to reinforce prejudice towards Roma and Travellers in the online

discussion task (Chapter 3), as well as my research assistants during the initial coding of the hateful online comments (Chapter 4). In particular, whilst the support of the confederates predominantly involved their ease to express (potentially) unpleasant statements during the discussions, the support of the research assistants involved several meetings questioning them about their well-being during the coding process.

Furthermore, since the start of this research programme, online user dynamics have changed. For example, what was defined as frequent use of Social Networking Sites (SNSs) or Social Media Platforms (SMPs) in 2011 is no longer the same. Specifically, people are ever more connected through mobile devices and the like, which enables them to be constantly available and thus inevitably increases how often they use SNSs and SMPs. This is particularly relevant for the study discussed within Chapter 5, where frequent use of SNSs/SMPs covered ‘several times a day’ up to ‘constantly’. This is problematic insofar that there are big differences between these frequencies and what they represent. Future research should therefore investigate these frequencies more carefully. In line with this, advances in technology are growing ever faster, resulting in research forever trying to catch up with it. For example, whilst the set-up of the study described within Chapter 3 remains to be new to research (i.e. online influence exerted during live instant-chat group discussions), many mobile applications now allow users to chat/communicate instantly with others without the need of sitting in front of a computer screen. As such, the realism of some of the aspects of the set-up might already be outdated.

Finally, the nature of who is online and is at risk of being influenced has changed since the beginning of this research programme. In particular, according to the EU Kids online survey (2014), compared to 2010, children aged between 11 and 16 are now up to 20% more at risk of being exposed to hate messages and up to 12% more likely to be

exposed to cyberbullying. As such, future research not only faces the challenge of trying to keep up with technology itself, but also with its resulting changes in user dynamics.

6.4 Implications of Findings

Below, the main policy, practice and research implications of the current findings are discussed. These are in the areas of spreading and preventing online hatred and online victimisation and offender rehabilitation.

6.4.1 Spreading and Preventing Online Hatred

Understanding online influence (e.g. in terms of spreading hatred) in order to reduce the increasing number of hateful online expressions and thus online victimisation is high on the global political agenda (European Commission, 2015, 2016), as each government has the responsibility to protect its citizens from harm. Despite this, there is currently neither a global definition of hate crimes or online hate crimes, nor is there an agreement on how to treat hateful online content. As a result, there is much variation in how Internet Service Providers (ISPs) police the problem, including an overall reluctance to remove hateful content (Nag, 2017). A global definition of (online) hate could therefore guide ISPs on what content is to be removed. In addition, it would aid law enforcement agencies around the world in their criminal investigations by providing clarity and equal treatment concerning the treatment of illegal online content.¹

Along the same lines, despite increasing political pressure on ISPs to take down hateful online content (BBC News, 2017; Clague, 2017; Kahn-Harris, 2017), to date, there

¹ Please note that such global definition is not to be understood as censorship of online content, as this could result in displacing the issue of hatred from public websites (Rohlfing, 2014) thus making us blind to ‘what is out there’ (Douglas, 2007) – which, in turn, may prevent the development of effective counter strategies.

is little legislation to help address the issue. Instead, while ISPs argue that it is not their responsibility to remove hateful content politicians in the UK have accused some ISPs to even make profit from such content (BBC News, 2017). It seems therefore unsurprising that (intended) new legislation aims at ISPs finances (e.g. fines up to €50 million, Clague, 2017). The study described in Chapter 4 is also an example of ISP's reluctance to remove hateful online content, as despite the video clip and many of the responses breaching YouTube's (not legally binding) user agreement, the clip is still available on YouTube.

However, ISPs have started to develop new strategies to counter online hate speech. For example, some ISPs have recently introduced mechanisms which flag up and allow users to report hateful content (e.g. Hate Free). However, there are some conceptual issues with these mechanisms, such as their reliance on a degree of user diversity, similar to the study described in Chapter 4. Specifically, the mechanisms are based on the principle that the power of the crowd will police itself, whereby the crowd is seen as one big community. Whilst this might be the case for some public forums as seen in Chapter 4, the way that the Internet and many forms of online engagement are constructed (e.g. through means of algorithms and filtering) points more towards multiple disparate communities, in which users chose their community based on them holding similar views. As a result, users will hear views which are evermore tailored towards (and thus reinforce) their own understandings (as intended by the study described in Chapter 3; Sunstein, 2001). In addition, what might be socially desirable in one group, might not be acceptable or even be offensive in another group. For example, it might be desirable to express right wing views in a related forum and thus it is unlikely that users of that community would feel the need to report certain (perhaps hateful) content. In fact, the results from the study described in Chapter 3 and other existing research demonstrates that hate groups already use the principles of filtering to connect 'like-minded' individuals and groups and reinforce their

views (i.e. spread hatred; Angie et al., 2010; DeKoster & Houtman, 2009; Glaser, Dixit & Green, 2002; Green, Abelson & Garnett, 1999; Wojcieszak, 2010) and to recruit new members to their cause (Anti-Defamation League, 2005; Castle, 2013; Douglas, 2007; Gerstenfeld, Grant, & Chiang, 2003; Nirvana News, 2016).

In addition, these mechanisms do not take the aspect of preventing expressions of online hatred into account, as they do not address or explain the risks of hateful content to online users. Yet, the importance of prevention has already been acknowledged, both academically and politically. For example, Moghaddam (2005) argued that prevention is a long-term solution to terrorism.² In addition, in 2015, the UK government introduced the so called Prevent duty to provide guidance on how to counteract people from being drawn into terrorism. This strategy is predominantly aimed at front line staff (e.g. teachers, lecturers, health workers, the police, the prison and probation services) and points out the risks of radicalisation and how to support those at risk. It is also in line with a relatively new trend describes SNSs as a means to individuals' (self-) radicalisation (Sukhni, as cited by Laubner, 2016) which, in turn, is defined as "a phenomenon in which individuals become terrorists without joining an established radical group, although they may be influenced by its ideology and messages" (Citizendium, 2009, para 1). In particular, there has been a recent increase in so called 'lone-wolf' or 'home-grown' terrorist attacks, which has been attributed, at least in part, to the Internet (e.g. Anders Breivik, see: Mala, 2011;

² As mentioned previously, there is much overlap in the different hate, radicalism, terrorism and extremism concepts (e.g. they are all motivated by aspects of hatred). Yet, it is important to note that there are distinctions in their legal definitions and underlying motivations. As such, the examples given here are to be understood based on the overlap in hate-based motivation and not their legal definition and/or political motivation.

Michael Adebolajo & Michael Adebowale, see: Whitehead, 2014). Prevent also aims to address the issue of individuals being recruited and radicalised online to then leave Europe to fight the ‘Jihad’ in Syria or Iraq (e.g. Amira Abase, Kadiza Sultana & Shamima Begum, see: Sky News, 2016; Ece B., & Merve S., see: Lauterbach, 2015; Kesinovic & Selimovic, see: Killalea, 2015; Yusra Hussein, see: Arkell, 2014). Here, the specific concern is that individuals can be hard to identify, which, in turn, hinders effective interventions and counter-terrorism strategies.

In addition, the combination of Muslim terrorist groups and repeated terrorist attacks carried out by Muslims in recent years have led many to view Muslim communities as suspicious and even stir up hatred against them (Nelles, 2016). Unsurprisingly, there are Muslim communities who feel unfairly persecuted and view the governmental counter-terrorism strategy (i.e. Prevent) to be aimed predominantly at them. Such feelings may then lead to increased community tension and feelings fostering division, grievance and even hatred. In line with this, the UK Home Office currently classifies international terrorist threats from Da’esh the most serious compared to other threats (e.g. far-right attacks; Boora, 2017).

6.4.1.1 The Role of (Social) Identification in Shaping Hatred. As mentioned before, interactions between individuals and groups who hold similar views (i.e. are considered ‘like-minded’), can contribute to the spread of hatred (Angie et al., 2010; DeKoster & Houtman, 2009; Glaser et al., 2002; Green, et al., 1999, Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Sunstein, 2001, 2009; Wojcieszak, 2010). In particular, individuals who identify with views similar to those advocated by relevant in-group members (e.g. hate groups), may subsequently internalise these views (i.e. are influenced by them) and adopt them as part of their own social identity (Turner, 1982).

For example, Chapters 3 and 4 provide some evidence to suggest that participants who identify with a message's content readily accept such views whilst rejecting and opposing message content with which they do not identify. In particular, Chapter 3 illustrates that (resistance to) online influence depends, at least in part, on how the proposed message content fits with people's existing attitudes (i.e. whether it is supportive or unsupportive of them). These findings therefore sit within existing attitude research (e.g. Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Petty & Cacioppo, 1981) and fit the small body of research into the (social) influence of online hatred (e.g. Coffey & Woolworth, 2004; Lee & Leets, 2002; Steinfeldt et al., 2010; Sunstein, 2001; Wojcieszak, 2010).

In addition, Chapter 4 highlights the importance of social psychological processes (e.g. social identification) in accepting or rejecting hateful attitudes. Specifically, there were two types of strategies observed, which fit within existing online hate research. First, distancing oneself from others (here, racists), through invoking negative stereotypes and/or using hateful language, including dehumanisation, as a means to highlight and reinforce their status as an out-group (Erjavec & Kovačič, 2012; Faulkner & Bliuc, 2016; Moghaddam, 2005; Mols & Jetten, 2014; Steinfeldt et al., 2010). Second, creating negative social comparisons to out-groups (Angie et al., 2011; Coffey & Woolworth, 2004; Douglas et al., 2005). In addition, the noted dehumanisation of out-group members mirrors previous findings (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Doughty et al., 2014; Erjavec & Kovačič, 2012) and is thought to create a sense of in-group superiority over the out-group (Bliuc et al., 2012; Mumford et al., 2008), as well as justifying punitive treatments of out-group members (Moghaddam, 2005).

Moreover, as pointed out previously, observed categorical or dichotomous thinking (i.e. distinguishing between 'us' and 'them') to illustrate one's particular group

belongingness (here, racists or non-racists) with others sits within the broader social identity approach (e.g. social identity formation and confirmation; Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). It also matches existing findings concerning online hatred (e.g. Angie et al., 2011; Bliuc et al., 2012; Charteris-Black, 2006; Verkuyten, 2013). In line with this, categorical thinking (e.g. here, positive self-representation and negative other-presentation and claims of preferential State treatments for others but not for ‘us’) further serve to create a sense of relative deprivation, socio-economic insecurity and social identity threat, including threats to one’s national identity. These findings therefore further correspond with existing hate related findings (e.g. Bowling & Phillips, 2003; Coffey & Woolworth, 2004; Douglas, McGarty, Bliuc & Lala, 2005; Faulkner & Bliuc, 2016; Mols & Jetten, 2014; Verkuyten, 2013; Young, 1999) and the social identity perspective (e.g. Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner et al., 1987). Finally, the observed contestation of the national (here, British) identity also mirrors the social identity processes pointed out above. It also reflects previous work concerning online hatred, national identity and political discourse (e.g. Cohen, 1994; Condor, 1997, 2000; Faulkner & Bliuc, 2016; Hopkins & Murdoch, 1999; Sindic, 2008).

6.4.1.2 The Role of Anonymity in Expressions of Online Hatred. Online anonymity comprises two main features; namely visual anonymity (i.e. online users cannot physically see one another) and limited channel (i.e. communicating via text only; Joinson, 1999). In addition, online users can anonymise or reveal their identities through the use of user names and avatars (i.e. “an icon or figure representing a particular person in a video game or Internet forum”, Oxford Dictionary, n.d.).

It is widely accepted that anonymity plays an important role in people behaving anti-normatively online and that identifiability can prevent such behaviours (e.g. Coffey &

Woolworth, 2004; Joinson, 1999; Suler, 2004) and the findings of Chapter 5 fit within this existing research. As such, the recent media/legal emphasis on online hatred, including ISPs being pressured to reveal user IP addresses, publicised convictions for online hate offences and/or, governmental counter-hate strategies as well as revelations concerning global surveillance programmes in news headlines (e.g. National Security Agency) could result in online users no longer considering themselves as being completely anonymous online. In any case, the increase in online hate offences would suggest that reduced perceptions of online anonymity are not a sufficient deterrent to commit online hate offences.

The results of the study described in Chapter 5 demonstrate that, indeed, perceptions of online anonymity are low but they also highlight that the media and publicised governmental anti-hate strategies do not affect these perceptions. Therefore, these findings provide some clarity and, in turn, allow to examine other factors which might affect perceptions concerning online anonymity. After all, the more we know about online users' perceptions of online anonymity the more can be done to prevent online hate speech.

Overall, it is clear then that an understanding of the factors influencing the expressions of online hatred can aid its prevention. Therefore, future research into the prevention of online hatred should always consider the role of identifiability versus anonymity. It should also take the relationship between the communicator of hateful content and its recipients into account, as online anonymity during group interactions (e.g. discussions) has been found to increase group identification, which in turn, increases social influence (e.g. can shape hate-based attitudes; Postmes, Spears, Sakhel & deGroot, 2001).

Taken together, the findings from this thesis can be implemented to develop new and extent existing counter hate strategies (see below). For example, they highlight the

importance of attitude strengthening methods (e.g. inoculation treatments), context (e.g. whether the message is spread amongst the 'like-minded' or on an open forum where diverse opinions are expressed) and awareness (e.g. the consequences of online victimisation) to policy makers, practitioner and researchers. They further show how audiences with diverse opinions towards racism engaged with materials depicting racial hatred. Moreover, the evidence concerning the (resistance to the) spread of online hatred amongst a not exclusively hateful audience also provides some answers to the overall concern that the Internet generally facilitates the spread of hatred. In particular, the results of this thesis illustrate that many people resist online influence, thus indicating that the suggestion of SNSs generally facilitating the spread of online hatred might be too simplistic.

The findings of this thesis, concerning the spread/prevention of online hatred, could also benefit (the UK) government(s) and a range of organisations (e.g. charities and ISPs) in their development and implementation of counter-hate/prevention strategies. For example, it could aid the government in the implementation of the Prevent duty. In particular, Prevent predominantly focusses on the identification of individuals at risk of being drawn to terrorism before providing support to counter such leanings. The findings of this thesis then suggest that attitude strengthening methods and inoculation treatments could prevent individuals from being influenced by hate groups at an earlier stage of developing such views. These methods also have the advantage that they do not single out individuals and/or specific groups of people as they a) only point out the risks, strategies and potential consequences of hate groups and b) can be delivered to groups (e.g. pupils, students). As such, extending Prevent with attitude strengthening methods and inoculation treatments has the potential to reduce the feelings of unfair persecution felt by Muslim communities and thus reduce might help to reduce a degree of community tension.

In addition, similar methods could be adopted by ISPs to make their counter hate strategies more effective. In particular, ISPs who operate open forums (i.e. are likely to be visited by a diverse audience) could extend their existing methods to flag up hateful content by also providing information about the risks and consequences of such content (e.g. online influence and victimisation). Such methods might therefore not only help in the removal of hateful content but might also prevent online influence and raise awareness and empathy towards the victims of online hatred, which, in turn, might increase the reporting hate-based incidents (online and offline) to the police.

Moreover, the findings could be used by other organisations delivering intervention programmes to tackle hatred (e.g. Sophie Lancaster Foundation, True Vision). In particular, these organisations aim to raise awareness of prejudice and hate crimes and provide them with ways to challenge it. The results of this theses further enable them to raise awareness of the risks of online influence and means to challenge it. These results also permit these organisations to address and thus potentially prevent other related issues, such as online bullying.

6.4.2 Online Victimisation and Rehabilitation

In terms of the effects of online hatred, the findings of Chapter 4 highlighted some of the consequences that hate on the Internet can have for its victims. In particular, and although the study described in Chapter 4 did not specifically focus on online victimisation, there was some evidence of online victimisation, ranging from abusive, threatening and harassing comments to responses of distress, shock and upset. In addition, the findings of Chapter 5 indicate that many online users do not believe that online hatred can result in offline victimisation. Specifically, online users not only blamed the victims of online hatred for their abuse, they also did not hold the posters of hateful online content

responsible for victimising others online. In addition, Chapter 5 highlighted that online users opposed legal consequences for inciting online hatred.

The results of this thesis could therefore also aid offender rehabilitation and thus the prison and probation services. In particular, rehabilitation programmes, could highlight the consequences of online victimisation to the relevant offenders. Specifically raising offenders' awareness of the damaging effects of online victimisation on its victims may also raise their empathy towards their victims. Such increased empathy might, in turn, help them in their rehabilitation process and possibly prevent them from inciting online hatred in the future. However, it should be noted that such programmes also depend on offenders' motivation towards the offence. For example, raising awareness of the victim impact might increase empathy in offenders who were motivated by thrill, yet, those who were motivated by retaliation or in response to a perceived threat might not feel empathetic towards their victim(s).

6.3. Methodological Considerations

6.3.1 Social Desirability

As highlighted earlier (see Chapters 1, 2 & 3), online hate research, as well as research into prejudice, racism and explicit attitudes is often subject to social desirability effects. In particular, expressions of prejudice/racism are frequently perceived to be sensitive topics and societal norms tend to disallow expressions of particular attitudes (Crowne & Marlowe, 1960; Furnham, 1986; Gawronski, LeBel & Peters, 2007; Joinson, 1999). As a consequence, responses concerning these areas of research are at risk of being untruthful.

Therefore, in an attempt to generally reduce the risk of participants adhering to societal norms (i.e. social desirability), this thesis maintained anonymity for participants in all of the studies in which they were likely to have responded in line with such norms

(Chapters 2 & 3). This decision was strengthened particularly for the study discussed in Chapter 3, whereby participants admitted after they completed the study that they would have responded differently (i.e. with less prejudice) had they been identifiable during the study. Additionally, the analysis carried out in Chapter 4 highlighted the importance of online anonymity in expressing online hatred insofar as that when creators of hateful online content felt threatened with potential legal consequences, they removed such content, which suggests a reduced sense of online anonymity.

In addition, researchers trying to explore explicit expressions of prejudice and/or racism (e.g. expressions of hatred) have to carefully consider their limited choice of target groups/topics. For example, the history of some groups, known to prejudice/racism research (e.g. Black and Jewish people), tends to disallow overt expressions of negative attitudes/prejudice/racism against them (Franco & Maass, 1999). Consequently, these groups are less suitable for online hate research, which relies on explicit communication. This is particularly relevant to the studies discussed in Chapters 2 and 3, which involved measuring prejudice against specific target groups, namely Roma and Travellers. In particular, these groups were precisely chosen in an attempt to try and overcome the social desirability aspects identified in previous work, as explicit expressions of negative attitudes against Roma and Travellers are said to be largely acceptable in Europe (Franco & Maass, 1999; James, 2014; Spears & Tausch, 2012).

Moreover, people seem generally reluctant to participate in studies which address explicit hateful attitudes in the first place and often fail to complete such studies, resulting in either biased sampling or a non-response bias (Couper, 2000). These biases, in turn, can prevent the generalisability of findings, as observed in the study reported in Chapter 5. Consequently, the specific aims of studies investigating explicit attitudes, prejudice and/or

racism can often only be revealed after participants have completed the relevant experimental tasks (e.g. Chapter 3).

6.3.2 Ecological Validity of Experimental Study Designs

In the study described in Chapter 4, respondents did not know that their responses would be used to inform our understanding of online hatred. Yet, participants in the other studies discussed in this thesis (see Chapters 1, 2, 3 and 5) knew that they were participating in psychological research, albeit they did not always know the specific aims of the relevant study. As such, there is only one study which can ‘truly’ claim ecological validity. However, as the findings obtained from the other studies are still largely in line with predictions and results from existing research, it can be suggested that these studies did not suffer from a lack of ecological validity.

In addition, although the majority of participants used in the study described in Chapter 3 resisted online influence which may have ran contrary to their existing views, the study cannot account for how long these effects might last for. For example, it remains unknown if the repeated/prolonged exposure to such views would result in participants being persuaded by or become more resistant to them. Although ethically difficult, for reasons such as the danger of ‘creating’ prejudiced participants over a period of time, future research should consider ways in which the sustainability of participants’ resistance to online influence could be measured. Such results are also needed for the adaptation/development and administration of (long-term) effective intervention programmes.

Moreover, neither of the studies in which hate/prejudice was expressed (i.e. Chapters 3 & 4) assessed how firmly participants believed in their stated views and thus how entrenched these views were. As mentioned before, it is possible that some participants who responded to the YouTube clip (Chapter 4), expressed certain (e.g.

offensive) views to stir responses to get a thrill out of other people's responses. Such behaviour has been frequently observed in so called 'trolling'. In addition, it is possible that some participants felt that it might be expected of them (i.e. requirement of the task) to express strong views during an online discussion (Chapter 3). Yet, it was methodologically difficult to assess how strong participants' views reflected the actual strength of their opinions in these studies, given that one design was purely observational and the other was already complex and time consuming for participants. The assessment of such reflections should nevertheless be considered in future research.

6.4. Directions for Future Research

There are a number of ideas that could be examined in future studies to help contribute to the still relatively small body of literature on online hate. Most of these ideas have already been considered in the discussion sections of Chapters 3, 4 and 5. To summarise, this thesis suggests nine main directions for future research. First, future research should explore online influence for alternative (ethnic) groups and groups which might be targeted for other identifiable traits (e.g. disability, faith, gender, transgender, alternative sub-cultures). Second, future research should examine the longevity of online influence. Third, the effects of online hatred with reduced online anonymity (e.g. visibility/identifiability through means of webcams or actual names) should be explored. Fourth, future research should compare online influence between identifiable and anonymous groups. Fifth, online influence in relation to high group identification should be examined (e.g. exploring whether liking and/or prior relationships of participants increases online influence). Sixth, a comparison between one-to-one (i.e. minority) influence and group (majority) influence should be considered in future research. Moreover, the number of confederates exerting online influence should be varied and compared in future designs. Seventh, the effectiveness of inoculation treatments to counter

online influence should be tested in future work. Eighth, alternative communication contexts in which online hatred is expressed (e.g. message/discussion boards, popular Social Networking Sites, gaming) should be explored. Finally, future research should explore the perceived permissibility of online hatred when it is aimed directly at participants or their peers.

6.5 Conclusion

Existing research into online hatred has largely neglected to examine if and how social networking might shape the attitudes of ‘everyday’ online users towards others. Therefore, the central aims of this thesis were to examine whether online hatred could indeed shape online users’ attitudes by changing their levels of prejudice towards others, as well as to find out how online users would respond to online hatred and how permissible they perceive it to be. The findings of this thesis, which were obtained from a variety of related novel studies with methodological and analytical diversity, demonstrate that: (a) online hatred can shape (negative) attitudes of certain online users; (b) exposure to online hatred does not lead to its inevitable endorsement; (c) perceptions of the permissibility and acceptance of online hatred vary. Overall, these findings further imply that the best way to counter online hatred is through raising awareness amongst users of its associated dangers (i.e. influence) and consequences for its victims.

6.6 References

- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Erlbaum.
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology, 41*, 627-657. DOI: 10.1111/j.1559-1816.2011.00730.x
- Anti-Defamation League (2005). *Attitudes toward Jews in twelve European countries* (May 2005). Retrieved from www.adl.org:
http://www.adl.org/israel/Eur_Poll_Israel_May_2005.pdf
- Arendt, H. (1951). *The origins of totalitarianism* (2nd ed.). New York: World Publishing C., Meridian Books, 1958.
- Arkell, H. (2014, October 1). Could Yusra have been lured to Syria by terror dating website? Mother's agony as police reveal teenager was on site called 'jihad matchmaker'. *Mail Online*. Retrieved from
<http://www.dailymail.co.uk/news/article-2776343/My-heart-torn-Mother-British-teenager-feared-fled-Syria-jihadi-bride-makes-emotional-plea-come-home.html>
- Barnett, R. (2013). *Jews & Gypsies: Myths & reality*. Ruth Barnett.
- BBC News, (2017, March 14). Facebook, Twitter and google grilled by MP's over hate speech. *BBC News Online*. Retrieved from <http://www.bbc.co.uk/news/uk-39272261>

- Blackwood, L., Hopkins, N., & Reicher, S. (2012). Divided by a common language? Conceptualising identity, discrimination and alienation. In K. J. Jonas & T. A. Morton (Eds.), *Restoring civil societies: The psychology of intervention and engagement following crisis* (pp. 222-236). Chichester: John Wiley & Sons Inc.
- Blackwood, L., Hopkins, N., & Reicher, S. (2013). I know who I am, but who do they think I am? Muslim perspectives on encounters with airport authorities. *Ethnic and Racial Studies*, *36*, 1090-1108. DOI: <http://dx.doi.org/10.1080/01419870.2011.645845>
- Blackwood, L., Hopkins, N., & Reicher, S. (2015). Turning the gaze on 'us': The role of authorities in the alienation of minorities. *European Psychologist*, *18*, 245-252. DOI: 10.1027/1016-9040/a000151
- Bliuc, A. M., McGarty, C., Hartley, L. K., & Muntele Hendres, D. (2012). Manipulating national identity: The strategic use of rhetoric by supporters and opponents of the 'Cronulla riots' in Australia. *Ethnic and Racial Studies*, *35*, 2174-2194. DOI: 10.1080/01419870.2011.600768
- Bora, S. (2017). *Preventing the spread of radicalisation in the UK*. Proceedings from tackling radicalisation in the UK 2017 conference. London: United Kingdom.
- Bowling, B., & Phillips, C. (2003). Racist victimization in England and Wales. In D. Hawkins (Ed.), *Violent crime, assessing race & ethnic difference* (pp. 154-170). Cambridge: Cambridge University Press.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The content and context of social identity threat. In R. Spears, B. Doosje & N. Ellemers (Eds.), *Social identity: Context, commitment, content* (pp. 35-55). Oxford: Blackwell.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101. DOI: 10.1191/1478088706qp063oa

- Castle, C. (2013, October 27). YouTube awards tarnished by YouTube hate videos and jihadi recruitment. *Music, Technology, Policy*. Retrieved from <https://musictechpolicy.wordpress.com/2013/10/27/youtube-awards-tarnished-by-youtube-hate-videos-and-jihadi-recruitment/>
- Charteris-Black, J. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society*, 17, 563-581. <http://dx.doi.org/10.1177/09579265060666345>
- Citizendium (2009, December 13). *Self-radicalization/definition*. Retrieved from <http://en.citizendium.org/wiki/Self-radicalization/Definition>
- Clague, B. (2017, April 17). Internet users in Germany approve of proposal to police hate speech online. *E-Marketer*. Retrieved from <https://www.emarketer.com/Article/Internet-Users-Germany-Approve-of-Proposal-Police-Hate-Speech-Online/1015706>
- Coffey, B. & Woolworth, S. (2004). “Destroy the scum, and then neuter their families:” the web forum as a vehicle for community discourse? *The Social Science Journal*, 41(1), 1-14. DOI: 10.1016/j.soscij.2003.10.001
- Cohen, R. (1994). *Frontiers of Identity: The British and others*. London: Longman.
- Condor, S. (1997). Having history: A social psychological exploration of Anglo-British stereotypes. In C. C. Barfoot (Ed.), *Beyond pug's tour: National and ethnic stereotyping in theory and literature practice* (pp. 213-253). Amsterdam: Rodopi.
- Condor, S. (2000). ‘Pride and prejudice’: Identity management in English people’s talk about “this country”. *Discourse & Society*, 11, 175–205. DOI: 10.1177/0957926500011002003
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464–494. DOI: 10.1086/318641

- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. DOI: 10.1037/h0047358
- DeKoster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me'. *Information, Communication & Society, 11*, 1155-1176. DOI: 10.1080/13691180802266665
- Doughty, M., Lawson, S., Linehan, C., Rowland, D., & Bennett, L. (2014). Disinhibited abuse of othered communities by second-screening audiences. *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video* (pp. 55-62). ACM. <http://dx.doi.org/10.1145/2602299.2602311>
- Douglas, K. (2007). Psychology, discrimination and hate groups online. In A. Joinson, K. McKenna, T. Postmes and U. D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 155–163). New York: Oxford University Press.
- Douglas, K. M., McGarty, C., Bliuc, A. M., & Lala, G. (2005). Understanding cyberhate: Social competition and group creativity in online white supremacy groups. *Social Science Computer Review, 23*, 68-76. <http://dx.doi.org/10.1177/0894439304271538>
- Erjavec, K. & Poler Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites'. *Comments, Mass Communication and Society, 15*, 899-920. DOI: 10.1080/15205436.2011.619679
- EU Kids Online (2014). *Findings, methods, recommendations*. Retrieved from <https://lisedesignunit.com/EUKidsOnline/html5/index.html>

- European Commission (2015, December 3). EU Internet forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online. *New Europe*. Retrieved from <https://www.neweurope.eu/wires/eu-internet-forum-bringing-together-governments-europol-and-technology-companies-to-counter-terrorist-content-and-hate-speech-online/>
- European Commission (2016, September 26). United against hate speech on the Web: Where do we stand? – Speech by Commissioner Jourová at conference with German justice minister Maas. *New Europe*. Retrieved from <https://www.neweurope.eu/press-release/united-against-hate-speech-on-the-web-where-do-we-stand-speech-by-commissioner-jourova-at-conference-with-german-justice-minister-maas/>
- Faulkner, N. & Bliuc, A. M. (2016). “It’s okay to be racist”: Moral disengagement in online discussions of racist incidents in Australia. *Ethnic and Racial Studies*, *39*, 2545-2563. DOI: 10.1080/01419870.2016.1171370.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behaviour: The MODE model as an integrative framework. *Advances in experimental social psychology*, *23*, 75-109. DOI: 10.1016/S0065-2601(08)60318-4
- Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology*, *29*, 469-477. DOI: 10.1002/(SICI)1099-0992(199906)29:4<469::AID-EJSP938>3.0.CO;2-S
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*, 385-400. DOI: 10.1016/0191-8869(86)90014-0

- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193. DOI: 10.1111/j.1745-6916.2007.00036.x
- Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, 3, 29-44. DOI: 10.1111/j.1530-2415.2003.00013.x
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58, 177-193. DOI: 10.1111/1540-4560.00255
- Green, D. P., Abelson, R. P., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice, D. T. Miller (Eds.), *Cultural Divides*, (pp. 429-464). New York: Russell Sage Foundation.
- Heidensohn, F. (1989). *Crime and society*. London: Macmillan Education.
- Hopkins, N., & Murdoch, N. (1999) The role of the 'other' in national identity: Exploring the context-dependency of the national in-group stereotype. *Journal of Community and Applied Social Psychology*, 9, 321-338. [http://dx.doi.org/10.1002/\(SICI\)1099-1298\(199909/10\)9:5<321::AID-CASP510>3.0.CO;2-M](http://dx.doi.org/10.1002/(SICI)1099-1298(199909/10)9:5<321::AID-CASP510>3.0.CO;2-M)
- James, Z. (2014). Hate crimes against Gypsies, Travellers and Roma in Europe. In N. Hall, A. Corb, P. Giannasi & J. Grieve (Eds.), *The Routledge international handbook on hate crime* (pp. 237-248). London: Routledge.
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31, 433-438. DOI: 10.3758/BF03200723

- Kahn-Harris, K. (2017). Why won't YouTube and Google consider moderators to tackle online hate? *The Guardian*. Retrieved from <https://www.theguardian.com/commentinfree/2017/mar/16/youtube-google-hate-speech-moderators>
- Killalea, D. (2015, November, 2015). Samra Kesinovic, Sabina Selimovic: How Islamic State recruits teenagers. *News.com.au*. Retrieved from <http://www.news.com.au/world/middle-east/samra-kesinovic-sabina-selimovic-how-islamic-state-recruits-teenagers/news-story/9a0ab035c6352e2eb93c93b1733bd76e>
- Lange, P. G. (2007). *Commenting on comments: Investigating responses to antagonism on YouTube*. Paper presented at the Annual Conference of the Society for Applied Anthropology, Tampa, Florida. Retrieved from <https://www.yumpu.com/en/document/view/14888058/commenting-on-comments-investigating-responses-to-antagonism->
- Laubner, A. (2016, September 27). Praevention beginnt im Kopf- Islamwissenschaftlicher erlaeutert Wurzeln und Ursachen von Selbstradikalisierung. *Mindener Tageblatt*, 226, 8.
- Lauterbach, J. (2015, June 17). Eltern der verschwundenen IS-Mädchen verzweifelt. *Welt*. Retrieved from <https://www.welt.de/regionales/hamburg/article142627415/Eltern-der-verschwundenen-IS-Maedchen-verzweifelt.html>
- Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist*, 45, 927-957. DOI: 10.1177/0002764202045006003
- Levinson, D. J., & Sanford, R. N. (1944). A scale for the measurement of anti-Semitism. *The Journal of Psychology*, 17, 339-370. DOI: 10.1080/00223980.1944.9917200

- Lloyd, A. (1995). *Doubly deviant, doubly damned: Society's treatment of violent women*.
United Kingdom: Penguin Books.
- Mala, E. (2011, July 22). At least 80 dead in Norway shooting. *The New York Times*.
Retrieved from <http://www.nytimes.com/2011/07/23/world/europe/23oslo.html>
- McDevitt, J., Levin, J., & Bennett, S. (2002). Hate crime offenders: an expanded typology.
Journal of Social Issues, 58, 303-317. DOI: 10.1111/1540-4560.00262.
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration.
American Psychologist, 60, 161-169. DOI: 10.1037/0003-066X.60.2.161
- Mols, F., & Jetten, J. (2014). No guts, no glory: How framing the collective past paves the
way for anti-immigrant sentiments. *International Journal of Intercultural Relations*
43, 74-86. DOI: 10.1016/j.ijintrel.2014.08.014
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in*
Human Behavior, 26, 1536-1546. <http://dx.doi.org/10.1016/j.chb.2010.05.023>
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of*
Personality and Social Psychology, 12, 125-135. DOI: 10.1037/h0027568
- Mumford, M. D., Bedell-Avers, K. E., Hunter, S. T., Espejo, J., Eubanks, D., & Connelly,
M. S. (2008). Violence in ideological and non-ideological groups: A quantitative
analysis of qualitative data. *Journal of Applied Social Psychology, 38*, 1521-1561.
DOI: 10.1111/j.1559-1816.2008.00358.x
- Myers, D. G., & Bishop, G. D. (1970). Discussions effects on racial attitudes. *Science, 169*,
778-779. DOI: 0.1126/science.169.3947.778
- Nag, P. (2017, April 24). EU to harmonize tech companies' methods in dealing with hate
speech on the internet. *International Business Times*. Retrieved from
<http://www.ibtimes.sg/eu-harmonize-tech-companies-methods-dealing-hate-speech-internet-9488>

- Nelles, R. (2016, December 20). Commentary on Berlin attacks. *Spiegel Online*. Retrieved from <http://www.spiegel.de/international/germany/editorial-on-the-suspected-berlin-terror-attack-a-1126762.html>
- Nirvana News (2016, January 5). Trump used in terrorist recruitment video. *Nirvana News*. Retrieved from <https://www.youtube.com/watch?v=cv1aU3pG3BQ>
- O’Keefe, D. J. (2004). Trends and prospects in persuasion theory and research. In J. S. Seiter & R. H. Gass (Eds.), *Perspectives on persuasion, social influence and compliance guiding* (pp. 31-43). Boston, MA: Person Education, Inc.
- Oxford Dictionary (n.d.). *Definition of avatar*. Retrieved from <https://en.oxforddictionaries.com/definition/avatar>
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: William C. Brown.
- Petty, R. E., & Krosnick, J. A. (2014). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1-24). Hove, East Sussex: Psychology Press.
- Postmes, T., Spears, R., Sakhel, K., & deGroot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology, 27*, 1243-1254. DOI: 10.1177/01461672012710001
- Reicher, S., Haslam, S. A., & Rath, R. (2008). Making a virtue of evil: A five-step social identity model of the development of collective hate. *Social and Personality Psychology Compass, 2*, 1313–1344. DOI:10.1111/j.1751-9004.2008.00113.x
- Rohlfing, S. (2014). Hate on the internet, In N. Hall, A. Corb, P. Gianassi, & J. Grieve (Eds.), *International Handbook on Hate Crimes* (pp. 293-305). London: Routledge.

- Sindic, D. (2008). National identities: Are they declining? *Beyond Current Horizons: Technology, Children, Schools and Families*, 1-22. Retrieved from http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/final_sindic_nationalidentities_20081201.pdf
- Sky News (2016, August 12). What became of Bethnal Green girls lured to Syria to join IS? *Sky News*. Retrieved from <http://news.sky.com/story/what-became-of-bethnal-green-girls-lured-to-syria-to-join-is-10532801>
- Spears, R., & Tausch, N. (2012). Prejudice and intergroup relations. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An Introduction to Social Psychology* (5th ed. pp. 449-498). Chichester: Wiley.
- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano Jr., L. A., Benton, E. & Steinfeldt M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology, 16*, 362-371. DOI: 10.1037/a0018692
- Sunstein, C. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.
- Sunstein, C. (2009). *Going to extremes: How like minds unite and divide*. New York, NY: Oxford University Press.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*, 321-326. DOI: 10.1089/1094931041291295.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Pacific Grove, CA: Books/Cole.
- Turner J. C. (1982). Towards a cognitive redefinition of the social group. In H. Tajfel (Ed.), *Social identity and intergroup relations*, (pp. 15-40). Cambridge, UK: Cambridge University Press.

- Turner, J. C., Hogg, M. A., Oakes, P. J. Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A Self-Categorization Theory*. Oxford & New York: Blackwell.
- Verkuyten, M. (2013). Justifying discrimination against Muslim immigrants: Out-group ideology and the five-step social identity model. *British Journal of Social Psychology*, 52, 345-360. <http://dx.doi.org/10.1111/j.2044-8309.2011.02081.x>
- Whitehead, T. (2014, November 15). Lee Rigby killer: From quiet Christians to Islamist murderers. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/uknews/law-and-order/10528365/Lee-Rigby-killers-from-quiet-Christians-to-Islamist-murderers.html>
- Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media Society* 12, 637-655. DOI: 10.1177/1461444809342775
- Young, J. (1999). *The exclusive society: Social exclusion, crime and difference in late modernity*. London: Sage Publications.

Appendices

Appendix I

Levinson and Sanford's (1944) Anti-Semitism (A-S) Scale

“This is an investigation of general public opinion concerning Jewish people. The following are statements with which some people agree and other disagree.”

1. It would hurt the business of a large concern if it had too many Jewish employees.
2. The Jews should make sincere efforts to rid themselves of their conspicuous and irritating faults, if they really want to stop being persecuted.
3. In order to maintain a nice residential neighbourhood, it is best to prevent Jews from living in it.
4. War shows up the fact that the Jews are not patriotic nor willing to make sacrifices for their country.
5. The best way to eliminate the Communist menace in this country is to control the Jewish element which disguises it.
6. Jews seem to prefer the most luxurious, extravagant and sensual way of living.
7. Much resentment against Jews stems from their tendency to keep apart and to exclude Gentiles from Jewish social life.
8. Colleges should adopt a quota system by which they limit the number of Jews in fields which have too many Jews now.
9. The Jews must be considered as a bad influence on Christian culture and civilization.
10. In order to handle the Jewish problem, Gentiles must meet fire with fire and use the same ruthless tactics with the Jews that the Jews use with Gentiles.
11. The Jewish districts in most cities are results of the clannishness and stick-togetherness of Jews.
12. Jews may have moral standards that they apply in their dealing with each other, but with Christians they are unscrupulous, ruthless and undependable.
13. On the whole, the Jews have probably contributed less to American life than any other group.
14. One thing that hindered the Jews from establishing their own nation is the fact that they really have no culture of their own; instead they tend to copy the things that are important to the native citizens of whatever country they are in.
15. A step toward solving the Jewish problem would be to prevent Jews from getting into superior, profitable positions in society, for a while at least.
16. The true Christian can never forgive the Jews for their crucifixion of Christ.
17. Jews go too far in hiding their Jewishness, especially such extremes as changing their names, straightening noses and imitating Christian manners and customs.

18. It is not wise for a Christian to be seen too much with Jews, as he might be taken for a Jew, or be looked down upon by his Christian friends.
19. When Jews create large funds for educational or scientific research (Rosenwald, Heller, etc.) it is mainly a desire for fame and public notice rather than a sincere scientific interest.
20. There is something different and strange about Jews; one never knows what they are thinking or planning, not what makes them tick.
21. The Jewish problem is so general and deep that one often doubts that democratic methods can ever solve it.
22. A major fault of the Jews in their conceit, overbearing pride, and their idea that they are a chosen race.
23. One of the first steps to be taken in cleaning up the movies and generally improving the situation in Hollywood is to put an end to Jewish domination there.
24. There is little hope of correcting the racial defects on the Jews, since these defects are simply in their blood.
25. One big trouble with Jews is that they are never contented, but always try for the best jobs and the most money.
26. The trouble with letting Jews into a nice neighbourhood is that they gradually give it a typical Jewish atmosphere.
27. It is wrong for Jews and Gentiles to intermarry.
28. One trouble with Jewish business men is that they stick together and connive, so that a Gentile doesn't have a fair chance in competition.
29. No matter how Americanized a Jew may seem to be, there is always something basically Jewish underneath, a loyalty to Jewry and a manner that is never totally changed.
30. Jewish millionaires may do a certain amount to help their own people, but little of their money goes into worthwhile American causes.
31. Most hotels should deny admittance to Jews, as a general rule.
32. The Jew's first loyalty is to Jewry rather than to this country.
33. It is best that Jews should have their own fraternities and sororities, since they have their own particular interests and activities which they can best engage in together, just as Christians get along best in all-Christian fraternities.
34. Jewish power and control in money matters is far out of proportion to the number of Jews in the total population.
35. Jewish leaders should encourage Jews to be more inconspicuous, to keep out of professions and activities already over-crowded with Jews, and keep out of the public notice.
36. I can hardly imagine myself marrying a Jew.
37. The Jews should give up their un-Christian religion with all its strange customs (kosher diet, special holidays, etc.) and participate actively and sincerely in the Christian religion.

38. There is little doubt that Jewish pressure is largely responsible for the U.S. getting into war with Germany.
39. The Jews keep too much to themselves, instead of taking proper interest in community problems and good government.
40. Jews seem to have an aversion to plain hard work; they tend to be a parasitic element in society by finding easy, non-productive jobs.
41. It is sometimes all right to ban Jews from certain apartment houses.
42. Jews tend to remain a foreign element in American society, to preserve their old social standards and to resist the American way of life.
43. Districts containing many Jews always seem to be smelly, dirty, shabby and unattractive.
44. It would be to the best interest of all of the Jews would for their own nation and keep more to themselves.
45. There are too many Jews in the various Federal agencies and bureaus in Washington, and they have too much control over our national policies.
46. Anyone who employs many people should be careful not to hire a large percentage of Jews.
47. On general fault of Jews is their over-aggressiveness, a strong tendency always to display they Jewish looks, manners, and breeding.
48. There are few exceptions, but in general Jews are pretty much alike.
49. Jews should be more concerned with their personal appearance, and not be so dirty and smelly and unkempt.
50. There seems to be some revolutionary streak in the Jewish make-up as shown by the fact that there are so many Jewish Communities and agitators.
51. The Jews should not pry so much into Christian activities and organizations, not seek so much recognition and prestige from Christians.
52. Jews tend to lower the general standard of living by their willingness to do the most menial work and live under standards that are far below average.

Appendix II

Participant Consent and Instructions (Chapter 2 - Study 1a)

Consent

The aim of my study is to develop a scale to measure people's attitudes. In particular, I am interested in measuring negative attitudes towards specific groups of people. I do not mean to imply that you personally are prejudiced towards others or that being prejudiced against others is normal (or not normal). Instead, I want to establish whether all of the attitude items attached, which originally came from a longer scale published in 1944, are measuring the same thing. My aim is to shorten and improve the original scale for contemporary research investigating negative attitudes and prejudice. It is important that you are aware that this study is anonymous. The items attached contain a number of statements, each of which you might agree or disagree to varying amounts.

I understand that the study is anonymous and the only identifiable information I am required to provide is my gender and my age. My data are being collected as part of a PhD research project, supervised by Aldert Vrij.

My data are to be held confidentially. This means that any hardcopies of my data will be kept in a locked cabinet for a period of at least five years after any resulting publication and my electronic responses will be kept on a password protected computer. To both, only the researcher and the supervisory team will have access to.

Participation is entirely voluntary and I have the right to withdraw my participation any time and for any reason during the study, and immediately after participation, without penalty. However, because data is stored in a de-identified way, it will not be possible to withdraw my data after the date of participation.

I will be able to obtain general information about the results of this study by contacting Sarah Rohlfig after 31 December 2013 (please note that as participation is anonymous it will not be possible to access feedback on individual performance).

If you have any questions about the above, please ask the researcher, by emailing her under (Sarah.Rohlfig@port.ac.uk) or, if you are completing the survey on paper, by asking her personally before continuing with the study.

If you agree to the above, which I sincerely hope you do, please click the next button below which indicates your consent to take part in the study.

Instructions - Positively Associated Group

To complete this study, I need you to focus on group of people you have positive feelings for. So, for example these people could be linked together by ethnicity, or by religion, all by occupation, all by a pastime, or by a team they support! Or anything else that links them together that generally makes them people that you like or have positive associations with! Then with this group in mind, please rate each of the following statements.

Now, before we go any further, I like to start by clarifying a few important points:

- Some of the statements express strong opinions which some participants may you feel sensitive about. But please answer as honestly as you can.
- I am aware that not all of the statements are applicable to all groups of people. I have therefore added not applicable option for you. However, please use this sparingly, and **ONLY** use this option when the statement truly does not apply to the group of people you have in mind. Using the not applicable option too often really does ruin my data so please using only when no other option makes sense.
- Please also note that you were **NOT** have to reveal the group you are thinking of to anyone at any time!

Once you know which group you have in mind, please rate each statement by ticking the box which indicates the amount you agree or disagree (using the scale below each statement).

Please take a moment once you have completed the ratings and read through the debrief form. It's important to me that you know why I have asked you to spend your time filling in my form.

Finally, thank you so much for your help!

Instructions - Negatively Associated Group

To complete this study, I need you to focus on a group of people you have negative feelings/connotations for. So, for example these people could be linked together by ethnicity, or by religion, or by occupation, all by a pastime, all by a team they support! Or anything else that links together that generally makes them people that you do not like or have negative associations with! Then with this within mind, please read each of the following statements.

Now, before we go any further, I'd like to start by clarifying the a few important points:

- I completely understand that this may be a difficult task to some, as I cannot and will not assume everyone how is negative attitudes towards others. However, most of us have been in situations where we have had negative experiences with others, which make us dislike them, even temporary and for a very short period. If you are struggling to think of the particular group, think of the situation where you dislike the behaviour of others and use the people from that situation to write the statements.
- Some of the statements express strong opinions which some participants may feel sensitive about. But please answer as honestly as you can.
- I am also aware that not all of the statements are applicable to all groups of people. I have therefore added and not applicable option for you. However, please use this sparingly, and **ONLY** use this option when this statement truly does not apply to the group of people you have in mind. Using a not applicable option too often really does ruin my data, so please use it only when no other option makes sense.
- Please also note that you were **NOT** have to reveal the group you are thinking of to anyone at any time!

Once you know which group you having mind, please rate each statement by ticking the box which indicates the amount you agree or disagree (using the scale below which statement).

Please take a moment once you have completed the ratings and read through the debrief form, as this is an important part to finish the study informed and happy, and it's important to me let you know why I have asked you to spend your time filling in my form. Finally, thank you so much for your help!

Appendix III

List of 46 Adapted Attitude Statements (Chapter 2 – Study 1a)

1. It would hurt in the business of a large concern if it had too many _____ employees.
2. _____ should make sincere efforts to rid themselves of their conspicuous and irritating faults, if they really want to stop being persecuted.
3. In order to maintain a nice residential neighbourhood, it is best to prevent _____ from living in it.
4. War shows up the fact that the _____ are not patriotic, nor willing to make sacrifices for their country.
5. _____ seem to prefer the most luxurious, extravagant and sensual way of living.
6. Much resentment against _____ stems from their tendency to keep apart and exclude people from their social life.
7. Colleagues should adopt a quota system by which they limit the number of _____ in fields which have too many of them now.
8. _____ must be considered a bad influence in Christian culture and civilisation.
9. In order to handle the _____ problem, people must meet fire with fire and use the same ruthless tactics with them that they used with other people.
10. The _____ districts in most cities are the result of clannishness and stick-togetherness of them.
11. _____ may have moral standards that they apply in dealing with each other, but with others they are unscrupulous, ruthless and undependable.
12. On the whole, _____ have probably contributed less to the western life than any other group.
13. A step forward solving the _____ problem would be to prevent _____ from getting into superior, profitable positions in society, for a while at least.
14. _____ go too far in hiding their ways of living, especially imitating other manners and customs.
15. It is not wise for people to be seen too much with _____, as they might be taken for a _____, or be looked down upon by their non-_____ friends.
16. There is something different and strange about _____; one never knows what they are thinking or planning, nor what makes them tick.
17. The _____ problem is so general and deep that one often doubts that democratic methods can ever solve it.
18. A major fault of the _____ their conceit, overbearing pride, and their idea that they are special.

19. One of the first steps to be taken in cleaning up the movies and generally improving the situation in the media is to put an end to _____ domination there.
20. There is little hope of correcting the biological defects of the _____ since these defects are simply in their blood.
21. One big trouble with _____ is that they are never contented, but always try for the best jobs and the most money.
22. The trouble with letting _____ into a neighbourhood is that they gradually give it a typical _____ atmosphere.
23. It is wrong for _____ and non-_____ to intermarry.
24. One trouble with _____ businessmen is that they stick together and connive so that a non-_____ doesn't have a fair chance in competition.
25. No matter how adapted a _____ may seem to be, there is always something basically _____ underneath, a loyalty to their own beliefs and a manner that is never totally changed.
26. _____ millionaires may do a certain amount to help their own people, but little of their money goes into worthwhile national causes.
27. Most hotels should deny admittance to _____, as a general rule.
28. The _____ first loyalty is to themselves, rather than to their country.
29. It is best that _____ should have known societies and clubs since they have their own particular interests and activities which they can best engage in together, just as others get along best in their societies.
30. _____ power and control in money matters is far out of proportion to the number of _____ in the total population.
31. _____ leaders should encourage _____ to be more inconspicuous, to keep out of professions and activities already overcrowded with _____, and to keep out of the public notice.
32. I can hardly imagine myself marrying a _____.
33. The _____ keep too much to themselves, instead of taking the proper interest in community problems and good government.
34. _____ seem to have an aversion to plain hard work; they tend to be a parasitic element in society by finding easy, non-productive jobs.
35. It is sometimes all right to ban _____ from certain types of housing.
36. _____ tend to remain a foreign element in society, to preserve their own social standards and to resist the Western way of life.
37. Districts containing many _____ always seem to be smelly, dirty, shabby and unattractive.
38. It would be in the best interest of all if the _____ would keep more to themselves.

39. There are too many _____ in the various government agencies and offices in Whitehall, and they have too much control, over our national policies.
40. Anyone who employs many people should be careful not to hire a large percentage of _____.
41. One general fault of _____ is their over-aggressiveness, a strong tendency always to display their _____ looks, manners and breeding.
42. There are few exceptions, but in general _____ are pretty much alike.
43. _____ should be more concerned with their personal appearance, and not be so dirty and smelly and unkempt.
44. There seems to be a revolutionary streak in the _____ make up as shown by the fact that there are so many _____ agitators.
45. The _____ should not pry so much into Christian activities and organisations, nor seek so much recognition and prestige from Christians.
46. _____ tend to lower the general standard of living by their willingness to do the most menial work and to live under standards that are far below average.

Appendix IV

Participant Debrief (Chapter 2 - Study 1a)

Research on attitudes has a long history in psychology. Much of this research has focussed on positive or negative attitudes toward one particular group of people. Classic social psychology has suggested that attitudes towards others have been associated with stereotypes and prejudice which in turn and that the consequence of ordinary categorisation of others (Allport, 1954; Tajfel, 1981). Ehrlich (1973) even went as far as arguing that no one can escape the existing attitudes and stereotypes which have been assigned to major ethnic groups. Stereotypes are also being used in propaganda by several websites (for example, the Ku Klux Klan), which can result in reinforcing existing negative perceptions, and may even lead to negative behaviour towards others (Chau & Xu, 2007). Many people have some degree of stereotypes towards others, but this does not always result in negative actions against them. Another difficulty with stereotypes is however, that stereotypes are automatically activated and people are often not aware of these negative thoughts processes and perceptions, which makes them very resistant to change (Devine, 1989).

The aim of this study is to develop an attitude scale which allows us to measure attitudes towards several groups, instead of one particular group only. This study is based on an original attitude scale developed by Levinson and Sanford in 1944.

As part of my PhD, I am developing an adapted version of the original scale addressing more than one group, and with fewer items. To be able to do this, I firstly have to find out if all items in this scale are reliable and equally important. This is the reason for the study you participated in today. The results from today will enable me to develop this new scale and allow me to use it to measure attitudes towards several other groups. The new scale also has the advantage that, whilst gathering information about several groups, it will be less time consuming (for both, participants and researchers) than the original.

If you would like to know more about this research and the findings, please contact Sarah Rohlfing at sarah.rohlfing@port.ac.uk after the 31st December 2013 by which time I should have finished collecting data and conducting analysis.

If this research has raised any issues or concerns, these can be raised to the Chair of the Psychology Research Ethics Committee at:

Chair of Psychology Department Research Ethics Committee,
Department of Psychology
King Henry I Street,
Portsmouth,
Hampshire,
PO1 2DY

Your participation has contributed to further our knowledge in an under-researched area.

Many thanks for taking part in my research!

Appendix V
Participant Consent and Instructions (Chapter 2 – Study 1b)

Consent

This study is the second study to develop a new attitude scale to measure negative attitudes towards others. In particular, the statements below are taken from an existing prejudice measure (Levinson & Sanford, 1944), which addressed one particular group (i.e. Jewish people). In contemporary attitude research this group is no longer suitable, hence the need to develop one for a suitable group. The original measure was also much longer and therefore difficult to administer in certain settings (e.g. investigations into online hate).

I understand that the study is anonymous. This means that although you will be asked to sign this informed consent form and provide your age and gender, these data will be kept separate and therefore cannot be linked together. My data are being collected as part of a PhD research project, supervised by Aldert Vrij.

My data will be kept in a locked cabinet for a period of at least five years after any resulting publication.

My data are to be held confidentially and will be kept in a locked cabinet for a period of at least five years after any resulting publication. Only the research and supervisors will have access to them.

Participation is entirely voluntary and I have the right to withdraw my participation any time and for any reason during the study, and immediately after participation, without penalty. However, because data is stored in a de-identified way, it will not be possible to withdraw my data after the date of participation.

I will be able to obtain general information about the results of this study by contacting Sarah Rohlfing after 30th June 2014 (please note that as participation is anonymous it will not be possible to access feedback on individual performance).

If you have any questions about the above, please ask the researcher personally before continuing with the study.

If you agree to the above, which I sincerely hope you do, please sign and date below which indicates your consent to take part in the study.

.....
(Signature)

.....
(Date)

Instructions

To complete this study, please rate the statements below (i.e. circle the appropriate number) based on how accurate you find the stereotypes to be for Gypsies. In particular, please rate each of the statements based on its general applicability of the stereotype to Gypsies and **not** how you personally feel about the group. A 1 indicates the statement does not fit Gypsies at all, a 10 indicates that it fits them perfectly.

Please note that the ratings do not indicate that you are in agreement of the stereotypes towards Gypsies. They are merely to indicate the suitability of the items for this particular group. Please also note that your participation is anonymous (you only need to provide your age and gender) and that the rating scales are displayed on both sides of the pages.

Appendix VI

Participant Debrief (Chapter 2 – Study 1b)

Research on attitudes has a long history in psychology. Much of this research has focussed on positive or negative attitudes toward one particular group of people. Classic social psychology has suggested that attitudes towards others have been associated with stereotypes and prejudice which in turn and that the consequence of ordinary categorisation of others (Allport, 1954; Tajfel, 1981). Ehrlich (1973) even went as far as arguing that no one can escape the existing attitudes and stereotypes which have been assigned to major ethnic groups. Stereotypes are also being used in propaganda by several websites (for example, the Ku Klux Klan), which can result in reinforcing existing negative perceptions, and may even lead to negative behaviour towards others (Chau & Xu, 2007). Many people have some degree of stereotypes towards others, but this does not always result in negative actions against them. Another difficulty with stereotypes is however, that stereotypes are automatically activated and people are often not aware of these negative thoughts processes and perceptions, which makes them very resistant to change (Devine, 1989).

The aim of this study is to develop an attitude scale which allows us to measure attitudes towards Roma and Travellers as part of a bigger research programme. As you may have noticed, the statements you rated referred to the term Gypsy instead of the less offensive term Roma and Travellers. I deliberate chose this term because people tend to be more familiar with it and tend to see these groups as one. However, it should be noted that these groups are distinctive from each other and outside of this study are seen so. This study is based on an original attitude scale developed by Levinson and Sanford in 1944.

As part of my PhD, I am developing an adapted version of the original scale which addresses Gypsies. Part of this process involves to test whether the stereotypes of the statements we selected in the first study indeed suit Gypsies or not. This is the reason for the study you participated in today. The results from today will enable me to finalise this new scale and allows me to use it as part of another of my PhD studies. The new scale also has the advantage that it will be less time consuming (for both, participants and researchers) than the original.

If you would like to know more about this research and the findings, please contact Sarah Rohlfing at sarah.rohlfing@port.ac.uk after the 30th June 2012 by which time I should have finished collecting data and conducting analysis.

If this research has raised any issues or concerns, these can be raised to the Chair of the Psychology Research Ethics Committee at:

Chair of Psychology Department Research Ethics Committee,
Department of Psychology
King Henry I Street,
Portsmouth,
Hampshire,
PO1 2DY

Your participation has contributed to further our knowledge in an under-researched area.

Many thanks for taking part in my research!

Appendix VII

Participant Consent & Instructions Session One (Chapter 3 - Study 2)

Pre- consent information

What does participation in this study involve?

There are two parts to this study.

(1) First, you will be asked to complete a brief questionnaire. This looks at your general attitude towards different groups of people. There are no right or wrong answers to these questions and they can easily be answered by ticking ‘agree’ or ‘disagree’. This part of the study should take no longer than **10 minutes**.

(2) The second part of the study involves taking part in an anonymous online chat between you and 2 or 3 other ‘chatters’. Because this part requires this amount of other participants, I will contact you as soon as there are enough volunteers. The online chat will be completely anonymous (i.e., you will not meet the other ‘chatters’) and will involve the discussions of a specific topic. The actual discussion topic will be specified on the day/time of the chat –but be assured that it will be a topic for which no prior knowledge or experience is required!

The online chat will be followed by a few questions about your “chat experience” and some general questions about the discussion topic. All in all, **the second part of the study should take no longer than 45 minutes**.

Once you have completed both parts of the study, I will provide a full debriefing (i.e. tell you about the specific ins and outs of the study).

Finally, despite the rapid growth of internet technology, so far, we know very little about online interaction and how particular topics are discussed in online environments. So, your participation in this study contributes to our understanding of this new area of research.

Consent

My study is an investigation into online social interactions during instant messaging.

The questions below make up the first part of my study. The questionnaire will establish your knowledge and views on a few controversial topics before proceeding to the actual online chat discussion in part two. Although there are several topics covered in the questions below, not all of them will relate to the discussion you will participate in. Please be assured that your participation is anonymous. There will be no web cam in the second part. There will be no right or wrong answers to any of the questions in the questionnaire or the online chat discussion.

After you have completed the first questionnaire, I will contact you as soon as possible to arrange a suitable time for you to complete the second part of my study.

Please note there may take a few days before I contact you because I need to make sure that I recruit enough participants for your group.

- I understand that my responses are anonymous. This means that I will not be identifiable to any other participants I talk to during the discussion task. However, I understand that the set-up of the study (i.e. its completion over two sessions) requires the researcher to contact me to arrange for a suitable date/time. Yet, I am aware that once I completed the study, the researcher will delete any email correspondence and my responses will be stored in a de-identifiable way (i.e. the responses will only be linked to a uniquely assigned participant number). My data and are being collected as part of a PhD research project, supervised by Aldert Vrij.

- My data are to be held confidentially. Any hardcopies of my data will be kept in a locked cabinet for a period of at least five years after any resulting publication. My electronic responses will be kept on a password protected computer only the researcher and the supervisory team have access to.

- Participation is entirely voluntary and I have the right to withdraw my participation any time and for any reason during the study, and immediately after participation, without penalty. However, because data is stored in a de-identified way, it will not be possible to withdraw my data after the date of participation.

- I will be able to obtain general and individual information about the results of this study by contacting Sarah Rohlfing after 30th June 2015 stating my uniquely assigned participant number.

If you have any questions at this point, or during the questions, please do not hesitate to contact me (sarah.rohlfing@port.ac.uk) before continuing with the study.

- Finally, if you agree to the above, which I sincerely hope you do, please make a note of your unique participant number (below) and use the following link (which indicates your consent) to take you to the first part of my study.

Your **Participant Number** is ()

<https://www.surveymonkey.com/s/onlinedispartone>

Thank you very much for your participation and for your time!

Participant Instructions – Session One

Please rate each of the following items, using the scales provided below them and by ticking the number which best indicates how you currently feel about them. Please pay attention to each rating scale as these might differ throughout the questionnaire (for example, they change from being 7 rating points to 4).

Thank you very much for agreeing to take part in my study on online interaction. I very much appreciate your time!

Appendix VIII

Overall Participant Instructions Session Two (Chapter 3 - Study 2)

Many thanks for agreeing to complete the second part of my study today.

At (**provide TIME**) please follow the first link provided below (if clicking on it does not work, please copy and paste it into your web browser to access this part of the study).

Please do not start beforehand as I am trying to provide everyone with an equal opportunity. I am not investigating what people know before the chat, but instead only interested in natural interactions between people.

In this first part, there will be 4 questions, including stating your participant number, which is (**provide PPT no**). In this part of the study, you will also receive instructions of the entire study and get given the discussion topic including some information about it.

After you completed the first 4 questions, there will be a page stating "**AT THIS STAGE YOU MUST HAVE COMPLETED THE ONLINE CHAT ROOM INTERACTION BEFORE PROCEEDING WITH THE REST OF THE QUESTIONNAIRE**". This will act as a prompt to enter the chat room. In order to do so, please follow the second link provided below, or if this does not work, copy it into a new window/page of your web browser.

PLEASE DO NOT CLOSE THE BROWSER WINDOW WITH THE QUESTIONS, AS YOU WILL NEED TO GET BACK TO THAT PAGE AFTER YOU COMPLETED THE CHAT.

On the chat room page, it will ask you for a name or alias. Please enter **ONLY** your **participant number** into the field! You may also change the colour of your writing if you like. Once you hit enter, you will be in the chat room.

I will also be entering the chat room, but will only monitor that everything is going okay and that no one has any technical problems.

Once all participants have entered the chat room, I will give the okay to start with the discussion. **I WILL NOT PARTICIPATE IN ANY WAY THOUGH.**

At the end of the chat, I will only give the okay for you to go back to the questionnaire on the first browser window, as the questions from that stage on will refer to your chat you have just completed. Please also close the chat room at this stage.

Once you are back at the questionnaire, it will take you about another 10-15 minutes to complete it.

After completing the questionnaire, you have also completed your participation in my study.

Thank you very much again for your help and please let me know if there are any problems or questions at any time.

First link (pre-and post-task questions): <https://www.surveymonkey.com/s/QSPMZZR>

Second link (Chat room): <http://www.chatzy.com/21765797454746>

Session Two: Pre-Task Instructions and Topic Information

Please sit yourself down comfortably by the desk in front of the computer screen.

In a minute, I will ask you to use the computer and the appointed chat-room to discuss your opinions and views with other chat room users about the actions of Basildon Council in evicting the Gypsies from Dale Farm. If you ever need to introduce yourself to the other chat room users, please only state your participant number to ensure anonymity.

I understand that for some given topic may be sensitive, which is why I need to make sure that you are happy with everything before we can begin.

Please be reminded that you have the right to withdraw from the study at any time and that I am happy to answer any questions before and after you have participated.

It is important that you pay attention to everything that everyone says during your online chat, as you may be asked some questions about this including who said what afterwards.

Appendix IX

Fact Sheet- Dale Farm Eviction (Chapter 3 – Study 2)

Dale Farm is a plot of land in Essex, United Kingdom. The site was previously used as a scrap and breaker's yard. Up until October 2011, Dale Farm's residents included predominantly Irish Travelers and some Romani. Half of the site is a legal camp with planning permission for occupation as a Traveller site, whereas at the other half, despite being legally owned by the Travelers, has been established without authorisation and planning permission. The former Dale farm residents state that they applied for planning permissions for the site on several attempts, but the council denied this on each occasion. In October 2011, the High Court ruled the eviction of the illegal site. Dale Farm residents have been forcibly evicted. The legal battle between the Dale Farm residents and Essex's Basildon Council has taken 10 years, cost £18 million and during the eviction riot police had to be deployed. At its height, Dale Farm housed over 1000 people, making it the largest Traveller community in the UK. The Travelers have been offered social housing, but rejected it, because this arrangement does not agree with their existing culture. If the Dale Farm residents had been allowed to stay, planning laws would have to be revised throughout the UK.

Appendix X

Manipulation Statements (Chapter 3 – Study 2)

Aiming to Reinforce or Refute Stereotypes Towards Roma and Travellers

Reinforcing

1. “I cannot believe that Basildon Council took so long to evict those Gypsies and that their best explanation to be allowed to evict them was that the Gypsies did not have planning permission!”
2. “I could not build an extension to my house without planning permission, so why should they? And also, the cleaning up of the land is going to take a very long time and will cost the taxpayer an astonishing £18 million!”
3. “It may be their own land but what makes them so special that they can avoid playing by the rules and simply doing what they like?!”
4. “It is their own fault they were evicted for breaking the rules!”
5. “Gypsies do not pay taxes anyway! If they want to stay where they are, they should start contributing towards society for once!”
6. “Even if they have no place to go now, it is their own fault and maybe they should have thought about that before they made themselves such bad names everywhere!”
7. “The reputation of being thieves and uncivilised must be true, otherwise why does this reputation follow them literally everywhere they go?!”
8. “The Council even gave them alternative housing suggestions, so what are they moaning about?!”
9. “I would not want to live next to or in the same area as Gypsies!”
10. “TV programmes like big fat Gypsy wedding demonstrated a true picture of what they are really like!”
11. “They are Travellers; the clue is in the name! They are not meant to stay in one place anyway!”

Refuting

1. “I cannot believe that Basildon Council evicted the Gypsies just because they didn’t have planning permission!”
2. “They had absolutely no choice than build their homes without planning permission, as they did apply for planning permission at the council, but the council continuously rejected them!”
3. “They even built their own waste water systems there and I would therefore argue that they improved the area instead of it simply being a deserted scrap yard!”
4. “It is even the Gypsies’ land, they bought it off the scrap yard owner and so should they be allowed to do what they like?!”
5. “They have no place to go now and in addition to that, the taxpayer will end up paying £18 million. Why?”
6. “The Council’s alternative housing for them does not consider, not appreciate the Gypsy culture, traditions or ways of living. What would the general population do if they were forced to live in caravans all of a sudden with no choice?!”
7. “What about their children, who are settled into schools and will lose their friends?”
8. “I think it was the fact that they are Gypsies which made the Council want them to leave! What would they say or do if it was themselves in the situation of the Gypsies?!”
9. “Who can blame them for wanting to stay where they were and defend the only space they were allowed to stay, with the prospect of nowhere to go to and not being allowed to stay in most places for very long anyway?”
10. “I think it was also the old stereotypes of Gypsies being thieves and the general misunderstanding of the general public towards Gypsy culture, putting pressure on the council to get rid of them.”
11. “TV programmes like big fat Gypsy wedding probably did not help the Gypsies in counteracting existing stereotypes.
12. “The language used in the build-up to the eviction was no different to the language the Nazis used in the build-up to the Holocaust!”

Initial Neutral Statements

1. "I guess Basildon Council evicted the Gypsies as a matter of principle, because this legal debate had been going on for about 10 years now."
2. "Apparently, those Gypsies had no planning permissions to build those houses they lived in, but on the other hand there was the rumour that there were several planning applications prior to them building, but the council rejected them all."
3. "They improved the site they lived on and even built their own waste water systems, but now that they have been evicted, it will apparently cost the taxpayer £18 million to clean the site up, which at the current financial climate is clearly not helpful for anyone."
4. "It is the Gypsies' land; they bought it off the scrap yard owner a long time ago. Perhaps the council should have thought about the consequences that this sale may have on their community at the time and bought the land themselves."
5. "There was the argument that the evicted Gypsies have nowhere to go now, but has that not always been their tradition, to travel and not stay in one place for very long?"
6. "The Council made some alternative housing available for them, what the Gypsies make of this offer is up to them."
7. "It may have played a factor in the lead up to the eviction that they are Gypsies. There are some stereotypes about Gypsies, including them being accused of being thieves, uncivilised and unclean. These stereotypes may have been adding pressure from the community onto the council."
8. "Perhaps the Gypsies really wanted to stay where they were, otherwise they would not have invested all this time and money to keep the place and make it their home, but is it right then to defend this place with barriers and threats of violence?"
9. "It may also have a negative impact on Gypsy children's education and social support, including the loss of friends and them lacking to bond with other non-Gypsies when they are being forced to move on."
10. "TV programme like big fat Gypsy wedding could even have a negative impact on existing stereotypes about Gypsies."
11. "The council was obliged to evict them, because it is the law and the council does not make the law."

Appendix XI

Example Discussion Transcripts (Chapter 3 – Study 2)

Please note, that in the first example the confederate has been assigned participant number **290** and in the second example number **9** (responses of the confederate are marked in bold)

Discussion Group 68 – Aiming to Reinforce Stereotypes

Sarah Researcher started the chat 55 minutes ago

280 joined the chat 58 seconds ago

267 joined the chat 40 seconds ago

Sarah Researcher: Hi there, we are just waiting on the others before I will let you start

290 joined the chat

Sarah Researcher: We are just waiting in one more to join you, so I would say we give it a few more moments

Sarah Researcher: Okay I would say we give it until 18:40 (3more minutes) then you can start, even if that means without the last person.

Sarah Researcher: Okay please start when you are ready

290: Okay

267: Hello

290: **Hi everyone. So.. what did everyone think about the dale farm eviction then?**

280: i think they were right to evict them

267: I have very mixed views, although I'd probably air towards the councils side

290: **I agree, 280. I'd say it was justified**

267: However I do see that if your culture wants to live a certain way, you should have that choice and can therefore see why they didn't accept social housing, however in terms of the actual eviction I do think it was the right choice

290: **To be honest, I'm surprised it took so long**

284 joined the chat

290: **hmm yes. But i suppose if you're going to break the rules, then you deserve to be evicted. If i can't build an extension on my own house without planning permission then there's no way they should be able to build.**

267: Yes I do agree with you, and for the fact that we have to buy land to live on, so there's no reason they should be able to get land for free. if they want to live how they do, that's fine but we shouldn't be at a disadvantage because of it

284: I agree 290 as rules were broken so the eviction was necessary

280: i agree, they broke the law, they couldnt predict the outcome of it

290: If they are travellers anyway, isn't the clue in the name? what i don't understand is why did they want to build permanent housing?

290: The council gave them alternative housing suggestions too, which I thought was fair

290: Well...almost too fair!

284: I agree that they were given a fair alternative which they chose not too accept

280: exactly

267: But I think to an extent it could have been assumed they wouldn't accept as they obviously have a very strong culture, so was that just to appease them rather than to actually offer it to them? I do agree that the eviction was correct, I just believe that many would have biased views due to stereotyping of the culture

284: Yeah there is a negative stereotype associated with the culture but that shouldn't have been an issue when considering the eviction surely?

290: I suppose, 267, that there is a negative stereotype. But the fact that they broke the rules and then built anyway only adds fuel to the fire.

284: 🤔

280: i dont think its based on the stereotpye, they broke the rules, thats why they were evicted

284: Harsh

267: I mean biased opinions about the whole situation rather than the decision of eviction

290: But the question is posed as to why this reputation follows them everywhere they go? Surely there must be an element of truth to it?

290: I'll admit that I would not want to live next to or in the same area as gypsies again. This is based from experience though, and not a stereotype

284: Reputations can be built on the actions of individuals rather than a whole culture though

280: why not? what was your issue with them?

267: But how many people truely know about the traveller culture and what they believe in, why they travel etc because I know I don't! All I know is it's a culture that's been around for a long time yet phrases such as 'gypsy' are so badly related to - this is due to what we've heard/learnt from originally a small number of individuals and not what the whole culture is about

284: Yeah completely true

290: They left the land in an awful state, they were threatening towards the public and were a complete hassle with the council. This time they were not on their own land, though.

284: Even so you cannot assume that is the case for them all?

267: But 290, your experience has caused you to create a stereotype by saying you wouldn't want to live near gypsies again you're insinuating that you believe this would happen again with different groups of people which I do believe is a stereotype or at least a huge generalisation

290: I just think that if they are breaking rules, or not on their own land or creating a hassle in society then they should not be allowed to stay there. Of course I'm sure this is not the case with all travellers, but it's most certainly the case for what I have dealt with in the past.

290: And by hassle I mean threatening, ignoring council requests, etc

290: I appreciate that they have their own culture that they should be able to live out, but that's going to be very difficult in a society of tax payers who, therefore, contribute to the society as a whole.

267: Many people do the above (threats, ignoring council requests) and many go a step further and ignore court orders or cause serious, harmful crimes such as murder of GBH, but that doesn't mean the whole population do those things such as not the whole traveler community would do as you've said

284: Well I'm sure if I'd had personal experiences like that I'd take certain measures to avoid that again

267: Unless you're suggesting eradicating the whole traveler community you'll always get certain individuals who will do certain things that will keep up their bad name, however you could say that for the rest of us too. To get rid of all the people who do or could potentially cause us problems you'd have to eradicate the entire worlds population

290: Of course 267, I don't think that those people you mentioned are better than travellers, but it does mean that I have little sympathy for examples such as Dale Farm when they were given warning, which was ignored, and they were evicted! I'd have the same view of someone who was evicted from their house

267 joined the chat 38 seconds ago

284: Yeah I think for Dale Farm the warnings were clear, the alternative was presented and ignored and therefore eviction was the only option left

267: And as I said earlier in the chat, I do agree that the dale farm eviction was the correct course of action, it's just the stigma and aftermath I have problems with 😊

284: Aftermath?

290: It's using their culture as an excuse not to abide by the law that I have a problem with, that's all.

290: what do you think, 280?

267: By aftermath I just meant the ever heightened stereotyping etc

284: Oh right

280: i agree with 284

284: Why's that?

280: the warnings were clear, they knew they had to leave, they are the ones that created the hassle

290: **I agree**

284: Fair enough, although there still deserve a certain degree of sympathy in my opinion

290: **to the traveller population, perhaps, but to those at Dale Farm I have very little.**

280: maybe sympathy in the fact they were rejected of planning permission several times, but this doesnt mean they could just break the rules

267: I wouldn't necessarily agree sympathy, more just understanding

Sarah Researcher: Okay guys. Thank you very much for your great contribution tonight. Can I ask you to please leave the chat room now and return to the rest of your questionnaires?

290: **Could we not say the same to them though? Did they try to understand why it is that planning permission was not granted? The uproar it would have caused?**

290: Okay, thank you 😊

Discussion Group 58- Aiming to Refute Stereotypes

Sarah Researcher started the chat 59 minutes ago

59 joined the chat 2 minutes ago

Sarah Researcher: Hello there, we spare just waiting on the others and then I will let you start 😊

59 joined the chat 95 seconds ago

50 joined the chat 90 seconds ago

59: No problem 😊

9 joined the chat 51 seconds ago

50: OK

9 joined the chat

Sarah Researcher: we are just waiting on one more group member. If they do not join us is the next 3/4 minutes I will let you start without them so you will not need to hang about

9: Okay 😊

Sarah Researcher: okay guys, looks like the last person is not going to join us today.

Thank you very much for helping me today and please start when you are ready 😊

9: Alright. So, what did you guys make of the Dale Farm eviction? Know much about it in the past?

50: I don't know much.. just read about it now honestly..

9: Like when it actually happened

59: I didn't i heard about it on the news but not in depth

59: Does anyone know how long ago it was?

50: I think in the questionnaire said 2011

9: Same here, 59. I did study it a bit at college too.. must have been about 3 years ago

59: Ah I see , does anyone have an opinion on what should of happened?

50: since you two probably know more than me, would you mind telling me what is the controversy about? I was reading that part of the land had the rights to be occupied as travellers site, part didn't but was owned by them.. is this correct>

9: I heard they're living (pretty much) down the road from the site in a bad state :/

50: so is that the reason why they were evicted?

9: Well, one of the main debates is... should they have been evicted? They owned the land, but some of it was green belt so they couldn't build without permission

59: 50, thats all that ive read i'm afraid. I think it has cost way to much if they were not creating problems for the surrounding community?

50: yeah I got that 9..

59: or maybe they were, people didn't understand their culture?

59: cousing conflict?

50: If they were evicted, do they still own the land?

59: **causing

9: Well, i can't help but think the negative stereotypes played a role in their eviction

9: I mean, if it was any other sub-culture in the U.K., would they have been treated the same?

59: totally agree 9!

50: yes.. but it's also true that there are some differences in the culture which make it difficult living in harmony - not saying one or the other is right

9: 50.. well, that's part of it. I'm assuming not - but how can they just do that?

59: Personally I dont know much about the travelling culture but surely part is to no remain in the same site for a long time?

9: what do you mean, 50?

59: *not

50: I don't know.. I think that on one side if they shouldn't have been living there it's fair that something was done about. on the other hand, I think circumstances should have been considered. perhaps it was possible to change the status of the land so that it was legal living there

9: 59.. These guys were known as "gypsies", but I think the term "traveller" is slightly misleading for this case study. These guys built housing.. a large community.. even sewage systems (if i remember correctly from college!)

59: Surely it would have also cost less!

9: They weren't going anywhere

59: Ah i see thank you, 9!

59: meaning with the cost of court battles etc

50: 9 I think that when there are strong cultural differences it may be difficult to live close by, share spaces etc

50: 59 probably yes

9: **Ah, I see, 50. So do you guys think they should have been evicted? Or not? I'm more on the side of not**

59: I think if there were major implication to the pre existing communities then no, why fix things if it isn't broke?

9: **And I just googled it.. it cost the tax payer £18 million to clear up everything after the eviction - what a waste! and to worsen their condition of living too.. I think that's disgusting**

50: no I think that some friction is bound to happen, but I wouldn't say that this mean the travellers were to go. who says they should leave and not the british who didn't like them? what I mean is just that it us understandable to have difficulties when there are differences

59: Totally agree 9! Obviously, there are disagreements in shared culture, but nothing so major to waste £18 million!

50: 59 I agree with that, I wonder if it was really necessary. I mean, the fact that was a greenland was actually the only reason?

50: Do you think that the opinion of the public may have put pressure on the decision?

59: The reseason being its a green land, sure they had already built on it. Therefore it has gone past the point of preservation?

9: **Exactly, guys. That's why I can't help but think that negative stereotypes (especially from programmes like My Big Fat Gypsy Wedding) may have been a factor in this**

59: Yes, totally. Again referring to the stereotypes.

9: **They did build on the green belt land.. that was their error. But to evict them for it? Why not just knock it all down (the stuff that shouldnt have been there anyway)**

59: I think the minority who were fame hungry, have blurred preceptions of their true culture

59: I think that it is totally unfair, that one mistake has cost them so dearly

50: 9 what would be the difference between knocking things down and evicting them? they couldn't live there anyway and they must have paid for what they built

59: also surely the damage to the green land had already been done?

9: The difference being that the majority (including the young kids who went to local schools there) shouldn't suffer by the actions of the minority

9: A lot of buildings were OK to be there

9: I'm not too sure, 59

50: 59 I don't agree on considering it "just a mistake"... but yeah sure they could have just given a fine or such and asked them to pay for getting permission to build

59: but the actions of the minority, by building them would have probably been used by the majority therefore benefitting them?

50: 9 I don't get it.. are you in favour or against the eviction then?

9: Against it. I felt that a) yes, they should have been punished, but b) a lot less harshly

50: 9 I think I agree with you

59: Does anyone know now what the land is used for?

50: no

9: If we really do live in a society where everyone is equal.. then that should be the case

50: that would be interesting to know..

9: and yeah.. it's pretty much just land now

9: After the clean up

9: Like I said, the gypsies have gone down the road and pretty much live (effectively) in the slums

59: So they have disrupted families lives in where a fine or less harsh punishment could have been given

9: I just wonder.. what was really achieved by the council?

59: *disrupted

9: Yeah, pretty much

59: Agree 9! however this may have been done due to public demand

59: *however

9: This is why I hate to generalise groups of people

9: It's so easy, but I always try to avoid it

9: Have you guys had much personal experience with gypsies in the past? Or just from the media etc?

50: I guess it's a normal mechanism..

59: in any community or group you are going to get all different kinds of people, a generalisation cannot be made although it happens

50: I had met them when I was a kid in a camping area abroad

59: In my job I have met many, mostly pleasant

50: so where does the negative opinion about them come from?

59: media, reality tv programmes, also exaggerations of stories

59: Land that they have accompanied near me, was left in a terrible state

59: *state

9: I've had some fairly neutral experiences with them, too. Some good some bad. I think for some people, it's easy to point the finger of blame.. and this may largely be pointed at gypsies because they are different.. they have a different culture.. a difference way of doing things (SOMETIMES) haha.. not that I wanna generalise 😊

59: Agree 9!

59: although my old work was trashed because of their inter rivalry

50: I think the problem is that it's part of the human nature to feel like a group and keep outsiders... out. I may sound more equal if it didn't happen, but it does. I wonder if we should rather accept it and see how it is possible to accept differences.. I don't if it makes sense

Sarah Researcher: okay guys. Thank you so much for your great contribution to this discussion. Can you please return to your questionnaires please and leave the chat room?

Appendix XII

Participant Debrief (Chapter 3 – Study 2)

Many thanks for taking part in my research! I hope you have enjoyed participating.

Research on online hatred has largely focused on webpage content analyses and the persuasive effects of materials people are exposed to on hate websites. Social identity theory and persuasion have not been investigated together, yet both seem to have strong links to hate behaviour. Hence, this is the focus of this research. The effects of online group interactions on individuals is a further area that has been neglected in the literature and is therefore also being examined.

This research investigates how individuals are influenced by others and how someone may identify with a group when interacting online with other in-group members. There were three different types of groups, which were pre-grouped by the responses they gave in first session of this study. The criteria for the pre-grouping were the attitudes (pro, neutral, against) expressed towards a minority group, in this case Gypsies. Also, a confederate (third or fourth discussion group member) led the conversation either to refute or reinforce existing attitudes of all participants.

From existing attitude research, it is expected that participants whose opinions are neutral towards Gypsies will be most persuaded towards the direction of influence compared to the other participants. This is because their attitude strength is somewhat weaker/less made up than those who hold either pro or anti sympathetic attitudes towards Gypsies. In particular, participants whose attitudes were either against or for Gypsies are said to have already formed strong attitudes and used counter-arguments, which should stop them from being persuaded by counter-argumentative positions. They are therefore expected to remain at their initial standpoint.

The set-up of this type of research is new and one hopefully shed light on whether individuals are influenced by other group members in an online setting in a similar way as they may be in a face-to-face interaction.

In the vital attempt to get people honest opinions, I would like to ask you not to reveal the true aim of the study to any of your friends or colleagues in case they may take part in my research.

If you would like to know more about this research and the findings please contact me at Sarah.Rohlfing@port.ac.uk after the 31st August 2014, by which time I should have finished collecting data and started to conduct the analyses.

If this research has raised any issues or concerns these can be raised to the Chair of the Psychology Research Ethics Committee at:

Chair of Psychology Department Research Ethics Committee,
Department of Psychology
King Henry I Street,
Portsmouth,
Hampshire,
PO1 2DY

Alternatively, if you are a University of Portsmouth member of staff or student, you can contact the University of Portsmouth counselling service at:

The Nuffield Centre
St. Michael's Road
Portsmouth
Hampshire
PO1 2ED
02392 843157

Your participation has contributed further to our knowledge in and under researched area and might even contribute towards tackling the shoes of online hatred in real life. Many thanks!

Appendix XIII

Initial Coding Scheme – Content Analytic (Chapter 4 – Study 3)

1	Dark Gray (highlighted)	Generally hostile/insulting
2	Light Blue (highlighted)	Indirect violence/incitement to violence by others/wishing violence on Emma (i.e. someone should do XXX to her...)
3	Yellow (Highlighted)	Direct violence against individual/woman, (i.e. I would kill, hurt, punch etc.)
4	Blue	Violence against a 'group' (US citizens, Low-income, Blacks etc.)
5	Blue (highlighted)	General prejudice
6	Olive (highlighted)	Addressing Emma West directly
7	Light Grey (highlighted)	General racist jokes
8	Dark red (highlighted)	Low level of racism
9	Pink (highlighted)	Medium level of racism
10	Medium Green (highlighted with green writing)	High level of racism/ sub human (e.g. calling black people monkeys)
11	Dark Blue Text 2 lighter 40%	General sexual remarks, including, directed at Emma or about the content
12	Gold	General Threats
13	Green (highlighted)	Drug/mental health related remarks
14	Purple	Re-posting (later deleted most of those comments)
15	Red (highlighted)	Feelings of posters/commenters (i.e. shame, sadness, embarrassment etc.) Part of condemnation of Emma's actions/behaviour
16	Teal (highlighted)	Remarks about Emma's child
17	Light Green (highlighted)	Condemning Emma and other racists/ anti-racist
18	Light green	Generally Supporting Emma
19	Accent 2 red darker 25%	Wannabe scientific (social scientific explanations for/ against a race their traits, superiority etc.)
20	Orange accent 6 darker 25%	Posting Emma's details/information about her/investigation
21	Purple accent 4 darker 25%	Feeling sorry for Emma
22	Orange Accent 6 darker 50%	Black people cannot be racist and comments related to it (i.e. all races can/cannot be racist)
23	Red Accent 2 lighter 60 %	Remarks regarding trolls (i.e. identifying or accusing other users as trolls and suggesting to ignore or mark them as spam/report them)
24	Black, Text 1 lighter 50%	General Governmental/Policies/Legislation/historical remarks (blaming Government, socio-economic climate, UK colonial history & treatment during those times, border control, UN, EU etc.)
25	Light Blue	Sorry/apologetic for content of the video/Emma's actions/racism, stating not everyone is like it in the UK, part of condemnation code.
26	Yellow	Referral to celebs (John Terry's wife etc.)

27	Black (original colour)	General remarks, not fitting any category ¹
28	Red writing and black highlight	Remarks about education, or lack of (Emma's education, racists' education are assumed to be low)
	----- or	Manual insertion of lines to indicate where posters/commenters are able to engage directly with other posters/commenters in sequential and related posts (displayed at the start and end of sequence to each other)

¹ This code does not relate to the comments we deleted for their un-relatedness or abbreviated application. Instead this code refers to comments, which we did not fit into any other code or were unrelated to the research question/online hatred.

Appendix XIV

Participant Consent and Instructions (Chapter 5 – Study 4)

Consent

The purpose of this study is to explore whether attitudes towards anonymity on the Internet have changed over time. In addition, this study aims to determine what terms or phrases people may perceive to be hateful, in order to compare these with relevant current legislation.

If you decide to take part in this study, you will be required to complete a survey on only one occasion, which will last approximately 15 to 20 minutes. Yet, there are no time restrictions on how long you may take to complete it. To keep your anonymity, you can complete the survey online and will not need to attend any sessions in person. However, there will be some questions relating to your demographics such as your age, gender and your place of residence. Then, there will be a series of questions relating to how anonymous you perceive yourself to be on the Internet and you will be required to rate some provocative statements with respect to whether you think they are hateful or not and whether such statement should be illegal or not. Please note that you will be able to decline answering questions you do not wish to answer.

Your anonymous responses will be kept securely by the principal investigator. In particular, the data will be stored electronically on a password protected computer. The data may be presented to others at scientific meetings, or published as a project report, academic dissertation, scientific paper or book, or in future research studies approved by an appropriate ethics committee. Your responses will not be passed to anyone outside the study team without your expressed written permission. The exception to this will be any regulatory authority, who may have the legal right to access the data for the purpose of conducting an investigation in exceptional cases. The raw data (your responses) will be retained for five years following publication (In line with APA guidance). When it is no longer required, the data will be disposed of securely.

You can stop any test at any time, or withdraw from the study at any time before finishing the survey, without giving a reason if you do not wish to. However, please note that because your responses will be anonymous, it will not be possible for you to withdraw once you have completed the survey.

If you have a concern about any aspect of the study, you should speak to the principal investigator in the first instance if this is appropriate or to the supervisor (both detailed below). If the concern or complaint is not resolved by the principal investigator or supervisor, you should contact the head of department Dr Sherria Hoskins, Sherria.Hoskins@port.ac.uk. If the complaint remains unresolved, please contact the University complaints officer (0239284 3642, complaintsadvise@port.ac.uk).

Principal investigator: Sarah Rohlfig
Telephone: 02392846614
Email: Sarah.Rohlfig@port.ac.uk

Supervisor: Aldert Vrij
Telephone: 02392846319
Email: Aldert.Vrij@port.ac.uk

This research is being self-funded by the principal researcher (Sarah Rohlfig). None of the researchers of the study staff will receive any financial reward for conducting the study, other than the normal salary or bursary as an employee or student of the University.

This study has been scientifically and ethically reviewed and given favourable ethical opinion by the Science Faculty Ethics Committee.

Thank you very much for taking the time to share your views with me.

If you agree to the above, which I sincerely hope you do, please click the next button below which indicates your consent to take part in the study.

Participant Instructions (Per Survey Section)

First of all, I would like to learn a little bit about your background (whilst maintaining your anonymity of course!). After that, I would like to know a bit more about your use of different social networking sites. Could you please provide the following personal information?

In this section, I'd like to hear more about your own personal use of social media. For each of the following statements, please do use the response that best describes your use of social media. For example, for the statements referring to your membership of social media sites and your mobile phone usage in relation to your use of the sites, you can simply select yes or no. In addition, for the statements referring to the frequency of your social media use, you could answer by typing things like several times a day, daily, weekly, monthly, less than once a month, or never etc.

In this section, I am interested in your views on some strong statements. People differ in what they actually consider to be offensive or hateful, especially in the context of social media. I am interested in finding out your personal views and just how hateful you think the statement below.

The statements below are all actual statements that people have posted online over the last four years, which have been taken Word for word (or in fact, letter by letter) from social media websites. You may find some of them unpleasant or upsetting. So, with that in mind, I'd be grateful if you could write each one according to your view on how hateful you think it is. Then, I'd be interested to know whether you think making the statements should be legal (i.e. whether or not the person making that statement should or should not be allowed legally, to make that statement publicly) or not, so you need to provide a rating for that too.

So to recap, for each statement you will provide two ratings - how hateful you think of this and then whether or not it should be legal to say such a thing. However, before I ask you to make these ratings, I would like to learn how you personally define 'online hatred' (i.e. what does the phrase 'online hatred' mean to you?). Please use the box below to describe in your own words your understanding of the term online hatred.

Now we are going to the statements that I mentioned above. Remember you are giving to ratings for each statement, by selecting one response option for hatefulness and one for whether you think the statement should be criminal or not. As before, this response options are displayed below each statement.

No I'd like to learn more about your views on a range of issues relating to our online and non-virtual real lives. Below you will find different attitude statements. The first lot concerns the expressions of strong or offensive use either online or offline. The second lot concerns the question of who should be blamed for and/or held responsible for dealing with online hate. Here, I'd like you to tell me the extent to which you agree with each of these statements. Please choose one of the response options besides the statement.

Glossary of terms:

Online anonymity - No sharing of personal data and all use of Webcams

Face-to-face - In person interactions, does not include webcams

Introduction of Smartphones - Smartphones were first introduced in 2007 and widely accessible by 2008

Introduction of Facebook - Although founded in 2004, introduction here refers to Facebook's worldwide popularity and use, which was approximately in 2009

Popularity of Twitter - Twitter was founded in 2006, but here the popularity of Twitter refers to the time when it was ranked third of the top 10 social networking sites in 2012

You have now read the last section of this questionnaire (you will be pleased to know!). Here, I'd like to turn to the issue of unlimited on the Internet. So finally, I briefly like you to tell me about how anonymous you feel on the Internet and whether you think anonymity plays a role in the way you behave online. As you did in the last section, please use one of the response options beside each statement to let me know how much you agree or disagree with it.

Thank you very much for your views and your time I don't have anything else to ask you, but if you think I've missed anything out, or if you have any comments you would like to share with me, please do so in the text box below.

Appendix XV

Full List of Proposed Measures' Statements (Chapter 5)

Hatefulness and Criminality

Control statements:

- (1) *"British is nothing but a 1707 political construct";*
- (2) *"Why can't we all just stick TOGETHER? When will you people realise it is a COLOUR and colour is not synonymous with being British".*

Offensive statements:

- (1) *"You're skin color is not british so youre not british"*
- (2) *"To be British, one must have poise and refinement. Black people possess neither of those qualities – they're just a bunch of dumb, boorish, violent thugs"*
- (3) *"If they are black or mixed race they are African, not British"*
- (4) *"Im okay with everyone.... except muslims.. they do not know how to integrate and they always make a mess anywhere they go. and their fucking religion of Allahdin"*
- (5) *"BLACK PEOPLE ARE SUPERIOR TO WHITES. WE ARE MORE TALENTED, SMARTER AND MORE BEAUTIFUL. Whites are useless"*
- (6) *"Trouble is .. blacks and muslims in the 3rd world are lazy"*
- (7) *"She's a woman, why is she allowed to leave the kitchen?"*
- (8) *"I don't usually have a problem with blacks, just your gang culture but thats it. and Muslims are the main problem, they hate all non muslims, not just whites but black people as well. and whats wrong with mosques? err maybe because they brainwash people in to hateing other people and commiting Jihad, and they are just Dirty scum, their not even worth talking about."*
- (9) *"FUCKING blacks, indians, pakis and polish and all of you! Go back to your own countries! Dont come over here and nick the jobs for us YOUNGERS! Fucking idiots!!!?"*

Borderline criminal statements:

- (1) *"Why don't whites go to nigger and paki countries? Because whites are smarter and have better countries"*
- (2) *"Niggers ruin everything they touch. Looks like they ruined Great Britain too"*
- (3) *"So what if whites are descended from Africans? We're smarter, less prone to violence, and in every way superior. Black people are lesser evolved apemen"*
- (4) *"Whites have achieved everything in human history, niggers, pakis and Irish haven't done anything."*

Criminal statements:

- (1) *"Dont speak but kill all the non whites on tram fucking invaders"*
- (2) *"DEATH TO THE WHITE RACE. KILL ALL WIPE THEM OUT"*

- (3) *“I agree that racism is a huge problem, and I have an ingenious idea for how we can eradicate it. Simply kill all niggers, chinks, Jews, etc., and we won't have anyone to be racist against. Problem solved.”*

Effects of online hate

- (1) *“Online posts, regardless of their content, are no more than words and can't hurt anyone in the real-world.”*
- (2) *“Social divisions are not related to anything said online”.*
- (3) *“Online threats do not pose a threat to anyone in the real world”.*
- (4) *“It is okay to say nasty things online because words cannot hurt anyone in the real world”.*
- (5) *“Online hatred does not lead to offline hatred”.*
- (6) *“Racist online posts can lead to social divisions in real life”.*

Online versus offline hate

- (1) *“Regardless of whether hate crimes are committed online or offline, all hate crimes are equally wrong.”*
- (2) *“Offline hate is worse than online hate because it can involve physical confrontation”.*
- (3) *“Not knowing who is targeting you is worse than knowing who is targeting you.”*
- (4) *“The effects of offline hate crimes are worse than those of online hate crimes.”*
- (5) *“Being targeted in person is worse than being targeted online.”*
- (6) *“Face-to-face physical threats are worse than anonymous online threats.”*
- (7) *“Online bullying affects people more than face-to-face bullying.”*
- (8) *“Being bullied online is worse than being bullied face-to-face.”*
- (9) *“Online hatred is worse than offline hatred because the re-victimisation is ongoing.”*

Freedom of expression

- (1) *“No one should be allowed to say offensive things to hurt others online.”*
- (2) *“People should be allowed to say whatever they want regardless of whether this is online or offline.”*
- (3) *“People posting offensive things about others should be arrested.”*
- (4) *“Legislation should not dictate what people can or cannot say online.”*
- (5) *“You shouldn't be allowed to say things online that you wouldn't say to someone's face.”*
- (6) *“The police should investigate all hateful online posts.”*
- (7) *“The police should arrest people posting hateful things online.”*

Responsibility-accountability

Victim:

- (1) *“It is your own responsibility to stop being bullied online by simply turning off the machine.”*
- (2) *“If people don't like what they read online, they should just leave the site.”*

- (3) *“People who are bullied online should simply ignore the bullies.”*
- (4) *“People being bullied online have only got themselves to blame.”*

Internet Service Providers:

- (1) *“Internet Service Providers should be held responsible for investigating hateful online posts.”*
- (2) *“It is the Internet Service Providers’ responsibility to remove hateful content that has been posted on their websites.”*
- (3) *“Internet Service Providers are to blame for allowing offensive online posts on their websites.”*

Poster:

“It is not the poster’s responsibility to consider how others might perceive their posts.”

Police:

“It’s not up to the police to investigate people posting hateful things about others online.”

Online anonymity

- (1) *“The evolving of the Internet doesn’t influence how anonymous I am online.”*
- (2) *“I am more anonymous on social media web sites, such as Twitter, than in face-to-face interactions.”*
- (3) *“I feel as anonymous as I did before the media recently highlighted the increased number of legal consequences for people posting hateful things online.”*
- (4) *“The Internet doesn’t make me completely anonymous.”*
- (5) *“Regardless of the recent increase in media coverage of the legal consequences of hateful online posting in social media, I feel I remain anonymous online.”*
- (6) *“I’d post nasty things online even if I didn’t actually mean them because no one knows it is me.”*
- (7) *“I feel less anonymous now since the number of legal consequences for people posting hateful things online increased (as reported by the media).”*
- (8) *“I feel as anonymous online as I did before the popularity of Twitter.”*
- (9) *“Recent publicised consequences for people posting nasty things on social media make me feel less anonymous online.”*
- (10) *“Since the introduction of Twitter, I don’t feel as anonymous online.”*
- (11) *“Before the introduction of smartphones I felt more anonymous online.”*
- (12) *“Anonymity doesn’t determine what I say towards others online.”*
- (13) *“On the Internet, I feel completely anonymous.”*
- (14) *“As the Internet becomes more advanced, I become less anonymous.”*
- (15) *“I feel as anonymous online as I did before the introduction of Facebook.”*
- (16) *“I feel bolder about expressing my opinions online than I do in person.”*

- (17) *"I'd behave differently online if I knew I could be identified."*
- (18) *"I wouldn't say certain things online if my friends knew it was me in case they'd judge me."*
- (19) *"I feel more anonymous online if I use a username (alias) instead of my real name."*
- (20) *"I feel I can say things online that I would never say out loud."*
- (21) *"I wouldn't have posted certain things online, if I could be identified."*
- (22) *"I felt more anonymous online before Facebook."*
- (23) *"Current legal pressure on Internet service providers to reveal people's identities when they behave badly online makes me feel less anonymous online."*
- (24) *"I'd remove some things I previously posted online if I thought they could be traced back to me."*
- (25) *"The recent increase in media coverage of investigations into hateful online posting makes me now feel less anonymous online."*
- (26) *"Anonymity doesn't determine how I behave online."*

Appendix XVI

Participant Debrief (Chapter 5 – Study 4)

Thank you for taking part in my survey. Whilst there is no direct benefit to you for your participation, the purpose of the survey was to investigate whether the attitudes towards anonymity on the Internet have changed all the time and whether perceptions of online hate much current legislation. Psychological research in the area of online hatred is relatively new whereas the Internet continues to evolve rapidly. Therefore, the possible relationship between anonymity and online hatred has not been explored yet. Specifically, we are interested in whether people perceive themselves to be anonymous online, given that there is more transparency on the Internet now than before. In addition, whilst there is relevant legislation regulating online hatred, we are also interested in whether people's perceptions on online hatred actually match this legislation.

Given the nature of the survey, the posting of provocative statements was required in order to achieve its aims. Some of the statements use all reflect sentiments that could be subject to interest from the criminal justice system. The rationale was to be able to investigate whether people's perceptions of criminality match actual legislation governing hate speech. Therefore, whilst the statements were deliberately chosen for their provocative and illegal and/or offensive content, these do not reflect the views of the research team in any way!

We hope that by examining the relationship between perceptions of anonymity and online hatred we will gain a better understanding of the issues surrounding this controversial topic.

If you become distressed as a result of completing the survey staff and students can contact the University of Portsmouth counselling/well-being services on (023) 9284 3157 or via email wellbeing@port.ac.uk, or, if you are not a student or member of staff of the University of Portsmouth you can contact victim support on (0845) 303 0900, who will be able to put you in touch with an appropriate person to talk to.

Please get in touch with either myself if you have any further concerns or questions. My contact details are as follows:

Department of Psychology

King Henry Building

King Henry I Street

Portsmouth

Hampshire

PO1 2DY

Telephone (023) 9284 6614

Email: Sarah.Rohlfing@port.ac.uk

Thank you once again for agreeing to participate in the study!

Sarah

If the concern or complaint is not resolved by the principal investigator or supervisor, you should contact the head of department Dr Sherria Hoskins, Sherria.Hoskins@port.ac.uk
If the complaint remains and resolved, please contact the University complaints officer (0239 284 3642, complaintsadvise@port.ac.uk).

Appendix XVII

UPR16- Ethics Checklist



FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Postgraduate Research Student Handbook for more information)

Postgraduate Research Student (PGRS) Information		Student ID:	422745
PGRS Name:	Sarah Rohlfing		
Department:	Psychology	First Supervisor:	Prof. Aldert Vrij
Start Date: <small>(or progression date for Prof Doc students)</small>	October 2011		
Study Mode and Route:	Part-time	<input checked="" type="checkbox"/>	MPhil
	Full-time	<input type="checkbox"/>	PhD
			MD
			Professional Doctorate
Title of Thesis:	The Dark Side of the Web? The Role of Social Networking in Shaping Hatred		
Thesis Word Count: <small>(excluding ancillary data)</small>	56,071		
<p>If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study</p> <p>Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).</p>			
UKRIO Finished Research Checklist:			
<small>(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-are-the-code-of-practice-for-research/)</small>			
a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES	<input checked="" type="checkbox"/>	NO <input type="checkbox"/>
b) Have all contributions to knowledge been acknowledged?	YES	<input checked="" type="checkbox"/>	NO <input type="checkbox"/>
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES	<input checked="" type="checkbox"/>	NO <input type="checkbox"/>
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES	<input checked="" type="checkbox"/>	NO <input type="checkbox"/>
e) Does your research comply with all legal, ethical, and contractual requirements?	YES	<input checked="" type="checkbox"/>	NO <input type="checkbox"/>
Candidate Statement:			
I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)			
Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	SFEC-2014-101A		
If you have not submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:			
n/a			
Signed (PGRS):	Sarah Rohlfing	Date:	04/01/2017

Appendix XVIII

Favourable Ethical Opinion (Chapter 2)

4

Part C Information for Submissions to the Ethics Committee for
Full Review

Title of proposed research
Pilot Study! Do indiv. and group interactions in online chatrooms influence ppl's attitudes and their self-esteem?

Name of researcher(s):
SARAH ROCHFING

This Ethics submission is for (please tick one of the following two options):

1) Full review for the first time

2) Expedited review (please tick one box)

A revision in response to Ethics Committee feedback. Please attach an additional sheet that details your responses to the concerns listed previously, along with the original submission.

Modification of already approved project – attach full previously approved proposal with a list of modifications or changes on a separate sheet

Departmentally-funded summer bursaries (or equivalent)

Practicals (i.e. 1st or 2nd year undergraduate)

M.Sc. unit with short deadlines (N.B. This does **not** include the dissertation)

Checklist for expedited and full reviews: Check that each of the following documents is enclosed with this form:

(a) Written responses to the items 1-19,
(b) Recruitment information (e.g. letters to parents, information sheet, Participant Pool poster, if applicable),
(c) Informed Consent Form (required),
(d) Debriefing Form (required),
(e) All questionnaires / Interview schedule – (if applicable).

.....

Decision of Ethics Committee: *Clare De* **Date:** *10/1/12*

Favourable opinion

Favourable opinion with provision [make the changes indicated on the proposal – no need to re-submit].

Unfavourable opinion - consult with your supervisor, tutor and/or mentor to rectify or address the concerns noted on the proposal, then resubmit following the instructions below:

No opinion possible [see proposal for details]

N.B. Revised proposals should be submitted in the Coursework box, Floor 1 King Henry Building. Remember to tick the first box under 2 above, tick the front sheet (Expedited review) and include (i) the original submission, (ii) the revised proposal (including a new cover sheet), and (iii) a list of your responses to the feedback.

4

Appendix XIX

Favourable Ethical Opinion (Chapter 3)

4

Part C Information for Submissions to the Ethics Committee for Full Review

Title of proposed research Online social interactions between group members
Name of researcher(s): SARAH ROHLFING

This Ethics submission is for (please tick one of the following two options):

- 1) Full review for the first time
- 2) Expedited review (please tick one box)
- A revision in response to Ethics Committee feedback. Please attach an additional sheet that details your responses to the concerns listed previously, along with the original submission.
 - Modification of already approved project – attach full previously approved proposal with a list of modifications or changes on a separate sheet
 - Departmentally-funded summer bursaries (or equivalent)
 - Practicals (i.e. 1st or 2nd year undergraduate)
 - M.Sc. unit with short deadlines (N.B. This does **not** include the dissertation)

Checklist for expedited and full reviews: Check that **each** of the following documents is enclosed with this form:

- (a) Written responses to the items 1-19,
- (b) Recruitment information (e.g. letters to parents, information sheet, Participant Pool poster, if applicable),
- (c) Informed Consent Form (required),
- (d) Debriefing Form (required),
- (e) All questionnaires / Interview schedule – (if applicable).

Decision of Ethics Committee: K. Jobson **Date:** 3rd April 2012

Favourable opinion

- Favourable opinion with provision [make the changes indicated on the proposal – no need to re-submit].
- Unfavourable opinion - consult with your supervisor, tutor and/or mentor to rectify or address the concerns noted on the proposal, then resubmit following the instructions below:

No opinion possible [see proposal for details]

N.B. Revised proposals should be submitted in the Coursework box, Floor 1 King Henry Building. Remember to tick the first box under 2 above, tick the front sheet (Expedited review) and include (i) **the original submission**, (ii) **the revised proposal (including a new cover sheet)**, and (iii) **a list of your responses to the feedback**.

4

Appendix XX

Favourable Ethical Opinion (Chapter 4)



Faculty of Science
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

Date 14/1/13

FAVOURABLE OPINION WITH MINOR AMENDMENTS

Proposal Title: A Content Analysis of comments made towards a YouTube video of a Racist Woman on the London Subway in November 2011

Dear Sarah,

Thank you for submitting your protocol for ethical review. The proposal has been on SFEC Moodle for two weeks for light touch review.

Your responses have been reviewed and I am pleased to inform you that your application has been given a favourable opinion subject to minor amendments/conditions (see attached docs) by the Science Faculty Ethics Committee.

You can go ahead with your research once you have provided me with your revised SFEC proposal so I can put the final version on Moodle.

Good luck with the study,

Best wishes,

A handwritten signature in black ink that reads 'Claire Nee'.

Dr Claire Nee
Psychology rep, Science Faculty Ethics Committee

CC -
Dr Chris Markham – Chair of SFEC
Dr Jim House – Vice Chair of SFEC
Holly Sawyer – Faculty Administrator

Appendix XXI

Favourable Ethical Opinion (Chapter 5)



Science Faculty Ethics Committee
Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

Ms Sarah Rohlfing
Department of Psychology
University of Portsmouth

T: 023 9284 3379
ethics-sci@port.ac.uk

sarah.rohlfing@port.ac.uk

24 June 2015

FAVOURABLE OPINION – SFEC 2014-101A

Protocol Title: Keyboard Warriors: Online anonymity and hatred

SFEC Code: SFEC 2014-101A

Date Submitted: 19 May 2015

Date reviewed: 20 May to 24 June 2015

Thank you for submitting your protocol amendment to the Science Faculty Ethics Committee (SFEC) for ethical review, in accordance with current procedures¹.

I am pleased to inform you that following review, your application has been given a favourable opinion by SFEC.

Please notify us in the future of any substantial amendments that may be required to this study, by making an application for protocol amendment².

SFEC wishes you well with your study³.

A handwritten signature in black ink that reads 'Jim House'.

Dr Jim House
Vice-Chair Science Faculty Ethics Committee
Review Chair

Information:

Supervisory Team: Professor Aldert Vrij (aldert.vrij@port.ac.uk)
Stefanie Sonnenberg (stefanie.Sonnenberg@port.ac.uk)
Samantha Mann (samantha.mann@port.ac.uk)

Holly Shawyer - Faculty Administrator

¹ Procedures for Ethical Review, Science Faculty Ethics Committee, University of Portsmouth, October 2012 (to be updated).

² Using the SFEC protocol amendment form.

³ If you would like to offer any feedback on the SFEC process please email ethics-sci@port.ac.uk to be forwarded to the Chair.

Appendix XXII

“Who is Really British Anyway?” A Thematic Analysis of Responses to Online Hate

Materials- Publication Version (Cyberpsychology)

“Who is really British anyway?”: A thematic analysis of responses to online hate materials

[Sarah Rohlfig¹](#), [Stefanie Sonnenberg²](#)

^{1,2} Department of Psychology, University of Portsmouth, UK

Abstract

This article aims to add to the relatively small body of literature on online hatred. In particular, it focuses on the role social networking sites may play in the development of polarisation, by exploring how online users respond to explicit online hate materials. Specifically, this article discusses the ways in which a self-selected sample of YouTube users responded, via posting online comments, to a video clip in which a White female train passenger (called Emma) could be seen to racially abuse other passengers. Thematic analysis of the YouTube comments identified four main themes: (1) Making Sense of Emma, which encapsulated posters' attempts to find explanations for Emma's behaviour; (2) Meeting Hatred with Hatred, which described posters' attempts to oppose Emma's racism by means of resorting to aggressive, hateful language; (3) Us versus Them, which encapsulated posters' tendencies to categorise themselves and other posters into in- and outgroups, based on their particular stance on racism; (4) Contesting Britishness, which expressed posters' attempts to articulate (and contest) what it means to be British. Whilst the current analysis provides some evidence that hateful web content can fuel aggressive and hateful responses, many of the comments analysed here emphasised common group membership, alongside people's right to claim membership in a particular social category (i.e. Britishness). The current evidence, therefore, suggests that, at least in the specific context of this study, hateful web content may not necessarily lead to an automatic endorsement or escalation of hatred.

Keywords: Thematic analysis; YouTube; online hatred; national identity; racism

Introduction

On the 21st of March 2016, 3,3 billion people worldwide were online (Internet Live Stats, 2016a). The relatively recent emergence of social networking sites (SNSs) has contributed noticeably to the growing numbers of Internet users. For example, Facebook had 1.6 billion users in March 2016 (Internet Live Stats, 2016b), and YouTube claimed, that in 2015, one billion users visited its page per month (YouTube, n.d.).

Despite their immense popularity, some SNSs have recently come under critical scrutiny. In Britain, for example, there is increasing government concern about the Internet's role in the incitement of hatred, including racial hatred. There have also been claims by some researchers, advocacy groups, Internet service providers (ISPs), and charities that websites promoting hatred may contribute to the development and exacerbation of radical attitudes and behaviours and, as such, foster terrorist sympathies (e.g. Foxman & Wolf, 2013; Perry, 2000; Stern, 2002). However, to date, claims regarding the persuasive power of the Internet remain largely based on anecdotal observations, common sense assumptions, and on government concerns regarding the protection of members of the public.

The existence of hateful materials online seems indisputable (e.g. it can be evidenced by certain key word searches). In addition, online hate materials, including propaganda video clips on websites, such as YouTube, are regularly found on computers seized from

terrorist suspects (e.g. the Boston Bombers, see: Robi, 2013; or the Lee Rigby murderers, see: Swinford, 2013). Consequently, lawmakers across the globe have begun to amend and develop legislation to account for the evolution of the Internet into a potential medium for stirring up and promoting hatred. In the United Kingdom alone, this legislation includes the Offences Against the Person Act, 1861, the Public Order Act 1986, the Malicious Communications Act, 1988, the Protection from Harassment Act 1997, the Crime and Disorder Act 1998, the Communications Act 2003, the Criminal Justice Act 2003 and the Racial and Religious Hatred Act 2006. Consequently, in the United Kingdom, several arrests and convictions have been made in relation to the publication and distribution of online materials with the intent to incite racial hatred (e.g. Colin White, see: BBC News, 2016; Ihjaz Ali, Kabir Ahmed, & Tazwan Javed, see: BBC News, 2012; Joshua Bonehill-Paine, see: Jewish News Online, 2015; Matthew Doyle, see: Burrows, 2016; Teenager, see: Crown Prosecution Service, 2010). Arrests have also been made in Australia, China, Dubai, Russia and Kazakhstan where several people have been convicted of inciting hate via social media, including ethnic hatred (e.g. Ai Takagi, see: AFP, 2016; Alkhanashvili, see: Kazakhstan Newline, 2015; Mukhtar, see: Radio Free Europe Radio Liberty, 2015; Pu Zhiquaing, see: Phillips, 2015; Tatyana Shevtsova-Valova, see: Latest.com, 2015), religious hatred (e.g. Yklas Kabduakasov, see: Spencer, 2015; UAE national, see: Agarib, 2016) and extremism (e.g. Vadim Tyumentsev, see: Reuters, 2015).

Academic research exploring online hatred has begun to emerge in several social sciences (mainly sociology and criminology). So far, this research has tended to have two main analytic foci, namely on providing (1) detailed descriptions of online hate materials or (2) characterisations of polarised individuals/groups. For example, research on online hate materials has focused on outlining the specific contents of hate-promoting websites and has provided insights into the use of propaganda to polarise and recruit new members to common ideological causes (e.g. Anti-Defamation League, 2005; Blazak, 2001; Bostdorff, 2004; Douglas, 2007; Douglas, McGarty, Bliuc, & Lala, 2005; Gerstenfeld, Grant, & Chiang, 2003; Levin, 2002; McNamee, Peterson, & Peña, 2010; Perry, 2000; Turpin-Petrosino, 2002; Zhou, Reid, Qin, Chen, & Lai, 2005). Research focusing on understanding the type of individual or group who produces hateful materials or hate speech online has tended to concentrate specifically on people who are already polarised, thus identifying not only some personality traits associated with hate speech but also some of the ways in which polarised individuals or groups share their ideology and deliberately use prejudice to create conflict between groups (e.g. Angie et al., 2011; Cammaerts, 2009; De Koster & Houtman, 2008; Doane, 2006; Erjavec & Kovačič, 2012; Ezekiel, 2002; Glaser, Dixit, & Green, 2002; Green, Abelson, & Garnett, 1999; Jacks & Adler, 2015; Wojcieszak, 2010).

However, thus far, research tapping into the broader *processes* underlying the development of online hatred remains limited – despite the reported increase in ‘hate blogging’, especially on SNSs, over time (e.g. Chau & Xu, 2007; Citron & Norton, 2011) and the recent surge in arrests and convictions for SNS-related hate offences. In particular, there is limited empirical evidence investigating how Internet users might respond to hateful online web content displayed on ordinary SNSs (i.e. YouTube, Facebook, Twitter). In particular, whilst one might expect to come across hateful/polarising content on certain websites (i.e. English Defence League, Stormfront), one might not expect to see such content during their visit to their familiar SNSs. As a result, we know little about how Internet users might respond to or be influenced by hateful online materials displayed on SNSs.

Many SNSs and online news web sites now allow, and even encourage, online users to express their views or responses to different web contents (including specific statements, pictures, video clips or TV programmes) via posting/commenting. These posts or comments are usually displayed publicly and, in turn, often seem to elicit further comments/posts from other users – who might agree or disagree with either the initial poster or with the web content in question. It seems therefore plausible that certain web

content could lead SNS users to respond to such content in a polarised manner. These polarised responses could, in turn, lead to (virtual) conflicts amongst SNS users.

In fact, such virtual conflicts can be observed in so-called online 'trolling', where certain individuals (i.e. 'trolls') intentionally aim to stir responses amongst users. Trolling has been attributed to online anonymity (Hardaker, 2010; Herring, Job-Sluder, Scheckler, & Barab, 2002), which, in turn, links to the so called online disinhibition effect (ODE; Suler, 2004). In particular, the ODE is thought to be the result of a lack of social constraints mediated, predominantly, by online anonymity (Joinson, 1998; Suler, 2004). Whilst not all behaviour associated with online disinhibition has to be negative (i.e. increased self-disclosure), as noted above, online disinhibition can encompass negative communication, such as offensive language, harsh criticisms, anger, hatred and threats (Suler, 2004). Suler (2004) referred to this form of the ODE as 'toxic' online disinhibition.

Accordingly, Coffey and Woolworth (2004) found that whilst people's responses to an alleged racially motivated murder reported in an anonymous online newspaper discussion board resulted in expressions of insults, racism, abuse and hatred (i.e. polarisation), no such responses were observed when the incident was discussed in a face-to-face meeting. They consequently attributed these polarised responses expressed online to online anonymity. In addition, Peddinti, Ross, and Cappos (2014) found that compared to identifiable Twitter users, anonymous users were less inhibited in relation to their Twitter activity. They also found that there are more anonymous Twitter users following sensitive topics, such as pornography, than identifiable ones.

Further, Doughty, Lawson, Linehan, Rowland, and Bennett (2014) found that controversial television broadcasts elicited many polarised (i.e. anti-social, abusive, negative, aggressive and hostile tweets - a phenomenon frequently referred to as 'flaming' (O'Sullivan & Flanagan, 2003; Suler, 2004; Turnage, 2008). In Doughty et al.'s (2014) study 'flaming', which is described as a negative aspect of the ODE, tended to be directed at the individuals or specific communities depicted in these broadcasts. In addition, Lange (2007) and Moor, Heuvelman, and Verleur (2010) investigated possible causes of 'flaming' behaviour on YouTube and found that people predominantly 'flamed' to express their disagreement with certain video or comment content. They also found that a small number of users 'flamed' for mere entertainment (Moor et al., 2010) and enjoyment (Lange, 2007). Lange (2007) and Moor et al.'s (2010) findings also indicate that, although 'flaming' seems prevalent on YouTube, it is not always perceived to be harmful, even by the recipients or 'targets' of 'flames'.

Moreover, Faulkner and Bliuc (2016) explored discursive strategies of racist and anti-racist supporters expressed on online news web sites. They found that moral disengagement is a very common strategy used by those supporting racism in order to justify or 'feel okay' about their views. They also found that those who opposed racist views/acts rarely used this strategy when expressing their views online. Given that expressions of racial attitudes are no longer socially acceptable (Crowne & Marlowe, 1960; Furnham, 1986; Gawronski, LeBel, & Peters, 2007; Joinson, 1998; Steinfeldt et al., 2010), it does not seem surprising that moral disengagement is predominantly observed in those supporting racist attitudes.

Current evidence therefore presents some strategies through which the expression of certain attitudes (i.e. racial) can be justified. In addition, it suggests that SNS content may not only facilitate discussions amongst online users but can also lead to attitude polarisation. However, to what extent does controversial web content succeed in influencing or changing people's actual attitudes or views?

Steinfeldt et al. (2010) found that online newspaper forums often provide misinformation about specific groups (i.e. American Indians, in their case) which, in turn, can reinforce existing negative stereotypes and thus facilitate the expression of racist attitudes. More specifically, Lee and Leets (2002) argue that White supremacist web content can

influence attitudes and views of polarised as well as 'neutral' adolescent online users. Furthermore, Rauch and Schanz (2013) found that, in the context of White supremacist online content, frequency of Facebook use influenced users in one of two ways. In particular, compared to infrequent Facebook users, frequent users were both more likely to agree with or reject racist messages (e.g. about White superiority), thus suggesting that those who frequently use Facebook are at higher risk of responding in a polarised manner to its content compared to those who infrequently use it.

In sum, the evidence outlined above suggests that the content of materials posted on online forums or SNSs can influence and polarise the expression of people's views as well as lead to online abuse— even in cases where the initial web content is not specifically problematic or hateful. Yet, the extent to which explicit online hate materials might elicit hateful responses in online users more generally remains unexplored. For example, given recent media coverage (e.g. Castle, 2013; McNab, 2015; Nirvana News, 2016; Poch, 2015; Risen, 2014) focusing on the use of YouTube and other social media sites (e.g. for promoting domestic violence or for terrorist propaganda and recruitment), it seems increasingly likely that people come across hateful content without necessarily seeking to do so. This then, raises the question of how SNS users in general might respond to *explicit* hate materials.

This study therefore aims to explore the following research question: *How do YouTube users respond to online materials that can be described as explicitly inciting hatred?*

Our research question here is deliberately broad in order to allow for a qualitative, bottom-up and inductively-driven analytic approach – which has been deemed the most appropriate research strategy for under-researched areas (Braun & Clarke, 2006). It is therefore important to stress that this study aims to offer an in-depth understanding of people's responses to a very specific type of online hate material (i.e. a video clip with racist content) within a particular SNS context (i.e. YouTube). In other words, this study seeks to provide a detailed description of the qualitative nature of people's online responses in a specific context – that is, it does not set out to make any claims about online responses to hateful materials in general. However, the authors hope that the qualitative findings outlined below will provide a springboard for further quantitative research in this field to evaluate the extent to which the key findings can be generalised.

Method

Materials

Data source. This analysis was based on comments made by users of the social networking site *YouTube* in response to a particular video clip uploaded on its website. The video clip depicted a White woman who racially abused other passengers. The woman, later identified as Emma West, was travelling with her toddler son on the London underground in October 2011. The scene, which lasted approximately five minutes, was filmed by a passenger on his/her mobile phone who subsequently uploaded it on 28th November 2011 as a video clip onto YouTube (for a detailed description of the clip's content, please see below). The clip prompted an unprecedented number of viewers to respond to its content via YouTube's commenting feature – and these responses provided the raw data for this study.

The nature of our choice of analytic materials was therefore somewhat opportunistic. In particular, when the video clip in question went viral online, the authors became aware of the analytic potential the accompanying comments presented for addressing the current research question. To the authors' knowledge, the video clip was the first of its kind (in the UK) to have elicited such large numbers of views and comments. Using this data source, therefore, had several methodological advantages. First, the voluntary exposure of YouTubers to the clip's content circumvented any ethical concerns an

experimental design (i.e. the experimental use of hateful stimuli) would have posed. Second, users' voluntary choice to comment on the clip's content was likely to render responses more ecologically valid than responses elicited in experimental settings. Third, unlike in many laboratory experiments, there was no researcher engagement with 'respondents', thus circumventing experimenter bias or social desirability effects.

The time-frame for data collection, in turn, was determined by the sheer volume of online comments made in response to the YouTube clip. Forty-eight hours after the clip's upload, the authors decided that the volume of comments was more than sufficient whilst still presenting a manageable analytic task. Also, the original video clip was removed at approximately the same time because the police started a criminal investigation of the woman depicted in the clip.

Consequently, the original raw data consisted of 71,295 comments. The actual content of the video clip is described below, followed by an outline of the different types of YouTubers' users' responses the clip elicited after its upload, as well as the legal consequences Emma West (i.e. the woman depicted in the video clip) faced for her behaviour.

Ethical considerations. The online comments (see above) that constituted the data source for this study were available within the public domain. As such, the responses were visible and freely accessible to anyone using YouTube. YouTube's terms of service explicitly state that, by uploading/posting content (including comments) onto the site, users automatically consent to this content being used (i.e. reproduced, distributed, prepared in derivative works) by third parties. However, in order to ensure that the YouTubers whose comments formed the basis of the analysis could not be identified further, all user names were anonymised.

The YouTube video clip. At the start of the video clip Emma West can be heard posing the rhetorical question "...*what has the UK come to...*" - which she then proceeds to answer herself by stating "...*a load of black people and a load of f***ing Polish*". Emma then verbally attacks several fellow passengers for "*not being English*", as well as telling them to "*go back to their own country*". Next, a Black woman asks Emma to refrain from swearing for the sake of the children present on the train. In response, Emma demands the Black woman state which country she is from, to which the Black woman responds that she, too, is English. Emma appears not to believe her. The clip continues by depicting Emma telling the same woman to "*go back*" to her "*own country*". Emma then attempts to specify which country the Black woman should "*go back*" to; she is, however, unable to articulate this correctly. In particular, she refers to an awkward combination of the two countries Nicaragua and Siberia, which in this specific combination sounds to other passengers like the racially offensive term N****R. This, in turn, can be seen to upset a young Black man behind Emma so much that he stands up, visibly agitated, in response to which two fellow passengers immediately get up and quickly calm him down; he then returns to his seat. Shortly after, a young White woman can be seen to rush towards the upset young Black man and embrace him. During this, Emma can be seen to argue with other female passengers calling out "*You ain't f***ing British! You are Black!*" Then a baby's cries can be heard, to which a seemingly angry White woman steps forward and shouts at Emma to "*shut up*" for waking up her baby. The woman can then be heard stating that she is English and asking Emma what she has to say to *her*. Whilst the argument between the two women continues, the video clip finishes.

Online YouTube responses. The video clip went viral after its upload and was viewed on YouTube more than 3 million times within the first 48 hours of its upload. Viewers' responses included comments making death threats towards Emma, her child, and other YouTubers, general threats of violence, as well as offensive and racist comments, which resulted in YouTube's removal of the clip from their website after approximately 48 hours. Since then, Emma West was arrested and charged with a racially aggravated public order offence. Emma pleaded guilty to racially aggravated harassment, alarm or

distress. Additionally, several copies of the video clip have re-emerged on YouTube, resulting in an estimated 11 million views to date, including repeat viewings.

Analytic Strategy

Data preparation. YouTube comments to the video clip were first imported into Microsoft Word and then formatted uniformly (Arial, size 12 with single line and one paragraph spacing between posts) to provide visual ease during the analytic process. This also aided initial familiarisation with the content. Next, comments that solely consisted of Computer-Mediated-Communication (CMC) abbreviation acronyms (Baron, 2004) such as *LOL*, and response tokens such as *yeah*, were deleted if the connection or relevance to other comments could not be easily established. Additionally, repeat comments made by the same users (i.e. displaying the same verbatim content) and comments made in a foreign language (30 in total) were deleted, leaving approximately 71.000 comments for analysis.

Data analysis. Based on the relative lack of existing research in this area, as well as the absence of an overall theoretical framework with which to account for how online users respond to online hate materials on SNSs, a bottom up, inductive approach to the data was taken. Overall, the thematic analytic process followed Allen, Bromely, Kuyken, and Sonnenberg's (2009) procedure. Generally, Thematic Analysis (TA) permits for a degree of epistemological and theoretical flexibility and is therefore not bound or limited to any specific pre-existing theoretical framework - yet, it does allow for the consideration of relevant theories during the analytic process (Braun & Clarke, 2006). This flexibility thus allows for the identification of themes at both the semantic (i.e. 'micro') and the latent (i.e. 'macro') level of analysis (Allen et al., 2009; Braun & Clarke, 2006). Given the broad inductive scope of our research question, both levels of analysis are drawn on here.

The thematic analytic process involved reading and re-reading the entire data set several times to become more familiar with the data and to identify themes related to the overall research question. A theme was deemed as such if, first, it tapped into relevant aspects concerning the overall research question and, second, if it represented a response-trend across the entire data set (Braun & Clarke, 2006). Reading and re-reading the data set continued throughout the formulation of themes, ensuring that the final themes remained data driven and that the chosen YouTube extracts (i.e. comments) accurately illustrated the corresponding theme/sub-theme. Themes were compared and were either separated further to show different nuances in meaning, or grouped together according to their overlap with one another (Allen et al., 2009). This process allowed for the development of an analytic hierarchy, by ordering the data into overarching themes and corresponding sub-themes (Allen et al., 2009). An overarching theme was defined as such if it appeared to be more inclusive (i.e. higher in the thematic hierarchy) and possessed greater analytic scope and explanatory strength than a sub-theme (Gleibs, Sonnenberg, & Haslam, 2014). Overall, the aim of the thematic coding process was to identify and organise emerging themes that were internally homogenous, externally heterogeneous and had explanatory power (Allen et al., 2009). Attention was also paid to prevalence which was determined by the frequency of responses that illustrated each theme (Braun & Clarke, 2006).

Results

Overall, the majority of responses condemned the content of the video clip. Analysis of YouTube users' responses to the video identified four overarching themes: (1) *Making Sense of Emma*, (2) *Meeting Hatred with Hatred*, (3) *Us versus Them*, and (4) *Contesting Britishness*, with the latter two being most prevalent. These main themes are examined in detail below under separate section headings. Each section includes a broad definition and a narrative description of the respective theme content as well as evidence from the actual online comments. Corresponding sub-themes will be presented as sub-headings.

It should be noted that here, as in other qualitative analyses (see Gleibs et al., 2014), is a degree of overlap between themes. For example, all themes reflect how posters, to some extent, attempted to position themselves (and other posters) in relation to racism. It should also be noted that many of the extracts below contain grammatical and spelling errors – which were only corrected in some cases to help clarify the meaning of these extracts.

Making Sense of Emma

Posters were obviously stirred by the content of the video clip and their responses often took a stance on racism. The video content seemed to evoke a range of negative emotions (i.e. expressions of distress) in several respondents. These emotions included shame (“*Such a rude, ignorant woman makes me ashamed to be British*”), disgust and embarrassment (“*You disgust me. You are the reason I am sometimes embarrassed to admit that I'm British.*”), upset (“*This video sickens me. She needs a reality check. I'm deeply upset by this footage.*”), and shock (“*I'm actually shocked! I can't believe what I have just watched?!*”; “*I am so shocked and appalled by this woman...*”).

Notably, respondents who expressed negative emotions condemned what they saw as Emma’s overt, racist behaviour. Many posters seemed to deal with what they had seen in the video clip by attempting to make sense of Emma and her behaviour. This *sense-making* focused on Emma’s personal characteristics, and particularly on attempting to attribute Emma's behaviour to internal, intra-psychic processes (i.e. alcohol and/or drug consumption):

OMG that woman is vile. And obviously either drunk or on drugs (judging from facial expressions and slurring). Luckily we don't all feel like her and she is one of a minority of people.. (Poster A)

This extract demonstrates how many respondents who disagreed with Emma’s racist behaviour, tended to attribute her behaviour to negative personal characteristics, which, in turn, enabled them to distance themselves from Emma. This distancing is evidenced by “*we don't all feel like her and she is one of a minority*” which, not only rejects racism but also positions the poster in opposition to Emma and other perceived racists and thus removes any basis for socially identifying with her. As will become clear later, the process of distancing also relates to another theme, namely *Us versus Them*.

The link between disapproval of Emma’s behaviour and attempts to make sense of it by focusing on Emma’s perceived negative personal and socio-economic characteristics becomes even clearer in the extract below. Specifically, here the poster attempts to make sense of Emma’s behaviour by explicitly referring to her perceived low intelligence, low education, and low socio-economic status. Moreover, here these personal characteristics (i.e. negative stereotypes) are seen to be associated directly with racist attitudes:

The fact she can't string her opinions without saying F*** says a lot more about her intelligence. would she happen to be on benefits and live in a council home by any chance? Its usually those people that moan about others but are happy to scrounge benefits and freebies of others who work in Britain and pay taxes like the people she insults (Poster B)

Here, there is an allusion to Emma’s racism as a direct consequence of her perceived low intelligence and low socio-economic status. In particular, these attributions seem to serve primarily to label Emma as a racist – in other words, as one of “*those people*” (i.e. an out-group member). Here, this strategy to position Emma as a racist (i.e. an out-

group member) further seems to serve as a means of making sense of Emma's behaviour.

In addition to explaining Emma's behaviour on the basis of a perceived lack of education, the following extract invokes a comparison between Emma and immigrants (i.e. immigrants "*have more skills*" and better work ethics than Emma). This comparison further serves as a way of discrediting Emma. The poster also tries to make sense of Emma's behaviour by linking it to a particular lifestyle (i.e. sexual transgression and work avoidance), which has been linked with poverty (Parisi, 1998):

This is what happens when u don't get an education that informs you of the realities of immigration. immigrants of all races have to work twice as hard to get jobs here and often have many more skills, in speaking several languages. I wonder what this ignorant fuck's skills are? probably shagging around and talking about how 'blacks' have ruined the country when she probably has never held down a job in her life! (Poster C)

Making sense of Emma's behaviour was also often associated with reference to Emma's role as a mother (i.e. "*a bad mother*" and "*a bad role model*"). In particular, responses frequently expressed concern for Emma's child, such as "*I feel sorry for her little boy*", "*she (i.e. Emma) used the child as a shield*", "*I hope social services will take her child away*", or "*poor child*". This, then, corresponds to broader strategies that have been observed in the media (Thomas, 1998) which aim to discredit single mothers through negative stereotyping and their portrayal as immoral and neglectful.

By referring to Emma and/or other perceived racist posters in hateful terms, comments frequently went beyond simply trying to make sense of Emma's behaviour, attempting to take a stance on racism:

Most of,if not all of the people backing this horrible cunt are unemployed dole spongers who use foreigners as an excuse for not getting a job.Waists of space racist pricks. most of the English football team are black. Show some respect (Poster D)

This, then, goes beyond trying to make sense of Emma: here, insulting and negatively stereotyping Emma and other racists serves to demonstrate the poster's opposition to their views. As will be seen, this use of hostile and aggressive language also links to the next theme (i.e. *Meeting Hatred with Hatred*). Also, the poster's reference here to the ethnic diversity of the English football team – a symbol of national pride – foreshadows the final theme *Contesting Britishness*.

Overall, *Making Sense of Emma* encapsulated respondents' attempts to understand and explain Emma. Specifically, respondents drew on a range of negative stereotypes both on a personal level (i.e. invoking alcohol/drug abuse, low intelligence, low education, 'bad mother/role model') and a social level (i.e. invoking a low socio-economic status) in order to account for Emma's behaviour and to distance themselves from (Emma's) racist views by positioning her as 'other' (i.e. an out-group member). This process frequently involved the use of strong, hateful language as the following extract illustrates:

Its funny how this silly lady thinks its 'the blacks' that are ruining her country when she is sitting on a tram with her son clearly on drugs and carrying on and the only person decent enough to tell her not to swear in front of kids is black..... I know who I would rather live next door two, the black lady and not the coked up whore. (Poster E)

Meeting Hatred with Hatred

This thick bitch would have got kicked the fuk out of by me the dik head in the back that takes his bag off why didnt he do anything.. Faggot bitch would get killed lil tart if i was there i would kill the bitch !!!! FUK YU DIE BITCH (Poster F)

This extract epitomises the essence of *Meeting Hatred with Hatred* as another overarching theme in this analysis. In particular, comments encapsulated by this theme depicted posters' tendency to strongly object to Emma's behaviour by means of demeaning and dehumanising her. Specifically, and somewhat paradoxically, many posters responded to Emma's behaviour by using hateful (including racist/abusive) and aggressive language, such as "*she has a mentality like Hitler*", "*she shouldn't be allowed to breed*", "*ugly ass throwback gobshite of a tredless bucketpussy usless eater oxygen thieving minus 80 I.Q*", "*lazy*", "*whore*", "*cunt*", "*chav*", "*slag*", "*anti-social, scum-sucking retard*", "*pikey*", "*albino monkey*", or "*honkey*". This use of hateful language in response to Emma's behaviour seems paradoxical in so far as it reflects an attempt to counter Emma's apparent racism and hatred with further expressions of hatred (hence *Meeting Hatred with Hatred*).

Along similar lines, posters frequently dehumanised Emma by referring to her as "*a piece of filth*", "*gutter trash*", "*White trash*", "*mosh mault*", "*monstrosity*", "*plankton*", "*a thing*" or a "*scum (bag)*". These terms invoke a perceived lack of humanity on Emma's part, thus placing her outside 'normal' human society and calling for her social exclusion (e.g. into "*the sewer/gutter/trash*", a place where society discards its unwanted waste). The dehumanisation of Emma thus represents another way in which posters attempted to distance themselves from her and the video's content.

In addition, posters who condemned Emma's behaviour called for her punishment, albeit on a symbolical level rather than a literal one. For example, posters referred to "*this lazy sponging cunt should be put down by a vet. Although animals are better looking that that scum.I sugest she goes back to where she came from the sewer*". As these examples illustrate, Emma's dehumanisation and calls for her punishment frequently went hand in hand – indeed, the latter seemed predicated on the former. These calls for punishment seemed to underscore further posters' apparent opposition to racism. Further calls for Emma's punishment included "*hitting her*", "*knocking her out*", "*beating her*", "*stabbing her*", "*slashing her throat*", "*deporting her*", "*locking her up in the sewer with the rats*", or to "*rot in hell*" – which, again, highlight posters' opposition to Emma's views and behaviour.

Overall, *Meeting Hatred with Hatred* encapsulated one particular method posters used to express their disagreement with the video clip's content. In other words, in their use of aggressive, abusive, threatening, and hateful language, posters attempted to distance themselves from Emma's expressions of racism. These attempts at distancing are also evident in the following theme.

Us versus Them

This prevalent theme describes how posters attempted to either distance themselves from or affiliate with Emma and her views by invoking a sense of "us" versus "them". This, in turn, echoes previous work focussing social identity processes (e.g. Blackwood, Hopkins, Reicher, 2012, 2013, 2015; Reicher, Haslam, & Rath, 2008; Tajfel & Turner, 1979; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). For example, posters condemning Emma tended to position themselves alongside other posters who shared their own seemingly anti-racist position – whilst, at the same time, distancing themselves from those who advocated or supported Emma's views (and vice versa). In particular, those posters who did not share Emma's views tended to refer to her and those who shared her views, as 'them' – in other words, as out-group members. Posters opposing racism also tended to refer to other like-minded YouTube users as *us*, that is, as in-group

members. *Us versus Them* consisted of two sub-themes that were labelled *Sense of Injustice* and *Sense of Deprivation*.

As already indicated, many posters condemned racist views and consequently distanced themselves from Emma:

@Participant X¹: [...] All racism is wrong. This woman is an embarrassment to where I come from. Hate engenders hate. Even the biggest, most intolerant racists must recognise exceptions to their misplaced and ignorant rules. Why not try giving everyone the benefit of the doubt? I don't want to be mistaken for someone like you, or this stupid woman on the video, so why should I assume anything about anyone else? (Poster G)

This poster distances him/herself from racism in two ways: first, on a social level, by calling "all racism wrong" and by labelling racists, as a group, as "intolerant" and "ignorant" and second, on a personal level, by describing Emma in negative terms, calling her an "embarrassment", "stupid", and "I don't want to be mistaken for (i.e. associated with) someone like you" (i.e. someone with racist views). Conversely, posters supporting Emma's views attempted to distance themselves from immigrants and immigration supporters:

@Participant X²: True, that was adorable how she stood up for her race and the bravery to say what everyone was thinking in front of adversity and to all the immigrant invaders who came to enforce their foreign oppressive laws and genocide [on] the land of the indigenous Whites. (Poster H)

This comment displays categorical thinking through its use of distancing pronouns, such as "their" *foreign oppressive laws*, implying that laws were not voted for or made by *us* (i.e. here, White people). Consequently, the poster here distinguishes between the in-group self (i.e. the 'us' supporting Emma's views) and 'the other' or out-group (i.e. those opposing Emma). Specifically, the above terminology depicts 'us' as good, while 'they' are bad. Here, the contrast between negative terms, such as "invaders", "oppressive", "genocide", "indigenous White", and positive terms, such as "adorable", "stood up" and "bravery" emphasises the distinction between the in-group and out-group (i.e. perceived racial differences). The term 'genocide' here also links to a perceived threat, ostensibly posed by immigrants, to posters' social identity (i.e. social identity threat; Branscombe, Ellemers, Spears, & Doosje, 1999; Reicher et al., 2008; Verkuyten, 2013) which will be discussed in more detail under *Contesting Britishness*.

Some comments in support of Emma not only created a sense of *Us versus Them* through use of in-/out-group defining pronouns, but also by displaying high levels of hostility towards out-group members, as illustrated in the following quote:

GOOD FOR HER! THEY CAN ALL FUCK OFF BACK TO WHERE THEY BELONG, AND THE TRAITORS WHO BROUGHT THEM TO THE U.K. MUST BE ARRESTED AND FUCKNG EXECUTED FOR TREASON!!! FUCK OFF! FUCK OFF! FUCK OFF!!! (Poster I)

This comment clearly demonstrates the poster's positive self-representation and negative other-representation, through out-group defining pronouns (i.e. *they/them*) and, in doing so, distances him/herself not only from 'foreigners' (i.e. 'they' do not 'belong' in the UK), but also from those people in the UK who support immigration (i.e. "traitors", who should be "executed for treason").

This hostility towards immigration supporters indicates a certain degree of conflict as the recipients of this hostility also belong to a potential *us* (i.e. the British). Here, British nationals supporting immigration are categorised as out-group members by those

opposing immigration which suggests that hostility expressed towards ostensible out-group members serves as a strategy to create inter-group conflict and justify action against them.

Overall then, *Us versus Them* tapped into broad social identity processes. That is, the theme and its two subordinate themes (see below) encapsulated posters' attempts to position themselves and other posters in line with the construction of in- and out-groups. As such, respondents' positioning frequently drew on the issue of immigration – which is exemplified by the following sub-themes, *Sense of Injustice* and *Sense of Deprivation*.

Sense of injustice. This sub-theme illustrates the way in which posters constructed a sense of *Us versus Them* by invoking immigration. In particular, those who voiced racist views tended to express their perceived sense of injustice by claiming that immigrants receive preferential treatment. Respondents who opposed racist views tended to argue that everyone “*born in Britain*” should be considered “*just as British as we* (i.e. White British) *are*” and therefore “*have the same rights to live in Britain*” (i.e. here, a sense of injustice consisted of denying immigrants their right to claim to be born in Britain as Britishness). These types of responses also related to notions of ‘Britishness’ and link to the final theme *Contesting Britishness*. There were also many posters who argued that immigration was an inevitable consequence of Britain's colonial history:

Hey lady - you don't like people from other countries or ethnic backgrounds in your country or calling them British or English - you guys started it! invading and looting all those other countries, staying in India, Pakistan, parts of Africa etc - now it's our turn!! PS. - rest of Britain who aren't racist like this stupid cow - we love you, we love your pubs, your beer & stouts, your football and your culture. Peace! (Poster J)

Again, here, a sense of *Us versus Them* is created through use of specific out-group-defining pronouns (i.e. *you* and *your* to address Emma and other British people), and in-group defining pronouns (i.e. *our* and *we* in relation to members of other countries and ethnic minorities). The poster also proclaims immigration an inevitable and direct consequence of British colonialism. In particular, his/her references to *invading* and *looting* express a sense of unfairness and injustice towards people from Britain's former colonies. Whilst this also suggests that s/he identifies as non-British, the poster also distinguishes between two different versions of British identity, namely a non-racist and racist British one, thereby differentiating between different types of British national group membership.

There were, however, also posters who endorsed Emma's stance on immigration:

[...] Islamic extremists do run round screaming “kill the infidel” and the police stand by and do absolutely nothing!. One law for black and another for white isn't working, any more than a housing or education policy that discriminates against established nationals, and merely fosters disharmony. (Poster K)

Although this poster does not directly categorise him/herself as a member of a specific in-group, this comment creates a sense of *Us versus Them* (i.e. Islamists extremists and Black people versus White people) through ethnic stereotyping. In particular, this post implies that all Muslims are Islamic extremists and not White – a misleading fusion which is likely to reinforce existing negative stereotypes. More specifically, the comment expresses discontent concerning the British establishment (i.e. laws and government policy), whereby the British criminal justice system is accused of unjustly differentiating between groups by “*discriminating against established nationals*”.

Whilst the poster in the above extract implicitly suggests that current laws/policies are unfair and contribute towards social division, other posters accused the British

government and criminal justice system explicitly of committing injustices against those who are 'really British'. Specifically, their sense of injustice referred to accusations of perceived persecution of 'real Brits' for 'minor crimes', such as "*expressing politically incorrect views*", whilst maintaining that 'severe' crimes, such as "*murder*", "*rape*", or "*burning poppies in the street*", committed by those 'non-British' (i.e. immigrants) "*remain unpunished*".

Thus, some responses to the video clip accused the British government of applying different rules to different people (i.e. favouring immigrants and discriminating against 'the British') - which directly related to posters' own attempts to differentiate between groups (i.e. *Us* versus *Them*). This sense of injustice (i.e. the claim that the establishment unjustly differentiates between groups) here also relates to the next sub-theme.

Sense of deprivation. The comments falling under this sub-theme conveyed posters' sense of feeling deprived. In particular, those holding racist and anti-immigration views tended to express a sense of deprivation via creating a sense of *Us versus Them* (i.e. here 'us' versus 'foreigners'). Some posters blamed 'foreigners' for their own financial insecurity and unemployment, including familiar phrases, such as "*they take our jobs*", "*they are prioritised for housing*" and "*they receive special benefits and free health care*", whilst "*we lose out to them*". Consequently, and as illustrated in the following extract, posters also expressed their sense of deprivation by drawing explicit group distinctions (i.e. racial out-grouping):

I wish more people would stand up for what they believe in we have a freedom of speech but somehow this has been withdrawn with this and why??? she has an opinion and is expressing it..... if people are offended then sorry they shouldn't watch it or be here to be offended against it..... we have got too many immigrants due to legislation..... and what a load of **** it is..... white population lose out on housing jobs and benefits because of foreigners..... how is that fair??? (Poster L)

The above reference to "*the white population losing out on housing, jobs, and benefits because of foreigners...*" not only suggests a sense of deprivation in general but, particularly, a sense of 'us' being deprived compared to 'them' (i.e. see fraternal relative deprivation; Runciman, 1966). For instance, 'our' (i.e. White British) socio-economic insecurity is here, blamed on 'them' (i.e. immigrants).

Overall, the theme *Us versus Them* reflected social identity processes as posters attempted to distinguish between different in- and out-groups (e.g. Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner et al., 1987). In particular, posters attempted to align themselves with other posters whose views they shared (*Us*) and to differentiate themselves from those whose views they opposed (*versus Them*). In doing so, comments frequently touched on notions of 'Britishness' - and thus specifically on the issue of who should be included in this social category and who constitutes *Us*. This then, relates to *Contesting Britishness*, the final theme, to which we will turn next.

Contesting Britishness

In the majority of posts, attempts were made to define what it means to be British by constructing and contesting different versions of British national identity. Here, posters did not simply try to position themselves in relation to Emma - instead, this theme indicates the extent to which respondents attempt to go beyond taking a specific stance on racism. In particular, this prevalent theme describes posters' ambivalence about the notion of 'Britishness' or national identity content, by capturing the ways in which comments contested the criteria by which 'Britishness' should appropriately be defined. This is perhaps unsurprising especially as the video clip showed Emma

challenging the British identity of a Black fellow passenger (i.e. referring to her as non-British). Also, given the diversity of YouTube users, one might expect variations in posters' references to national identity and sense of Britishness.

Whilst endorsement of racist views and an anti-immigration stance need not necessarily go hand in hand, there frequently appeared to be a degree of overlap between the two. For example, here, respondents who positively embraced immigration did not always seem to differentiate between racism on the one hand and anti-immigration on the other. They therefore frequently referred to those who rejected immigration as racists.

Overall, respondents' accounts of the criteria defining 'Britishness' ranged from being "multi-cultural", "tolerant", "accepting of others", "patriotic", "born in Britain", "White", to "holding a British passport", or rejecting the existence of 'true Britishness' altogether:

She doesn't have respect. No one is PURE english no more, not even the queen and the royal family is and I bet she isn't either. I'm mixed my father is english and my mum's thai, I was born here in the UK and I am a BRITISH. I follow every thing she does. .. I do the same as she, what just because My mum's from a different country? Just because these people are different colour means there not British.. HOW SHE WAS SAYING THIS IS MY BRITAIN.. NO BITCH THIS IS OUR BRITAIN. LIVE WITH IT! (Poster M)

Here, the poster clearly rejects Emma's version of 'Britishness' and, particularly, the view that anyone can be considered 'purely' British – including the ultimate symbol of 'Britishness', namely the Royal Family (ETHNOS, 2005). In particular, s/he rejects the idea that 'Britishness' can be determined by one's skin colour. In doing so, this poster not only distances him/herself from Emma's views, s/he also implicitly affirms immigrants' and non-White people's right to claim a British identity.

The following two sub-themes relate to the degree to which the boundaries of 'Britishness' are understood as inclusive of 'others'. These were *Britishness Under Threat*, in which multiculturalism was claimed to be either the 'undoing' of 'Britishness' or, conversely, as its very essence, and "*We are all Immigrants*", in which notions of 'Britishness' were explicitly linked to Britain's history.

Britishness under threat. The content of a number of posts seemed to express fears that Britain might be 'losing' its 'Britishness' due to immigration. In particular, these fears tended to relate to posters' perceived inability/restriction to express their 'Britishness' (i.e. not being allowed to "fly the George Cross on St. George's Day", or to "refer to Christmas"). In other words, here, 'Britishness' is seen to be under threat. Along the same lines, some posters expressed a sense of identity threat by references to "cultural genocide" or "cultural cleansing" in relation to immigration. Here, such references create a sense of social identity threat (Branscombe et al., 1999; Reicher et al., 2008; Verkuyten, 2013) amongst those supporting Emma whilst simultaneously invoking conflict with the perceived out-group (i.e. foreigners).

Interestingly, most posters who self-identified as English appeared to endorse a British national identity (rather than an English one). In particular, respondents who explicitly referred to their English heritage rarely voiced concerns about losing their *Englishness* – instead, when national identity was seen to be under threat this tended to be expressed in terms of the *British* losing their 'Britishness'. In addition, posters frequently differentiated between a British *we/this country* from a non-British *them/other countries* (Condor, 2000):

She has a valid point, albeit her language is really strong. The white, working classes in this country are becoming increasingly marginalised by the flood of immigrants from Africa, Asia and Eastern Europe. She's frustrated and angry. The present and preceding governments have failed her for years,

and all she sees on public transport are ugly, threatening black and brown faces. Welcome to Tony 'War Criminal' Blair's vision of Britain for the 21st century. (Poster N)

Here, the British government is reproached for allowing "*the flood of immigrants*" into Britain – the poster thus distances him/herself from the government's vision of Britain. This distancing is further emphasised by stating that the government "*have failed*" the "*White working classes*" which, together with references to "*ugly, threatening black and brown faces*", underscore the poster's sense of threat to his/her national identity. Also, the poster's use of the term "*marginalisation*" is interesting here. First, the term conveys a sense of (relative) deprivation for which s/he holds immigrants responsible (see also above). Second, the use of this particular term here expresses a sense of identity threat by implying that White people have become a minority in Britain.

However, many other posters rejected this notion of *Britishness Under Threat* altogether. These posters argued in favour of immigration, suggesting that it enhances and enriches 'Britishness'. This, in turn, was linked to calls for greater equal rights in terms of who should be allowed to call themselves British:

I'm British (white) and know I have no greater right to call myself British than those who became citizens today. I hope the ignorant minority of white nationalists will flow into our cultural past, but I doubt it. No society ever lives without ignorance, prejudice or racism completely. I would say I'm ashamed of this woman but I see no connection between myself and her just because of our race. Don't get frustrated by her ignorance, Britain's a multicultural society, and I'm proud of that. (Poster O)

Here, then, multiculturalism becomes the very essence of what it means to be British – which is in direct opposition to the views and behaviour depicted in the YouTube video. The poster clearly dis-identifies with Emma (i.e. "*see no connection with her*") and distances him/herself from White nationalists in general (i.e. by referring to them as "*ignorant*"). Rather than threatening British national identity, here multiculturalism is construed as defining 'Britishness'.

Other posters argued that throughout British history different cultural groups came to Britain and therefore shaped 'Britishness'. This not only gives people the right to call themselves British, it also undermines 'Britishness' as a fixed category. Some posters even suggested that "*We* (i.e. the British) *are all Immigrants*" – a claim which we will discuss next.

"We are all immigrants". Many YouTube comments explicitly and positively embraced immigration, especially given Britain's colonial history. For example, some posters responded to those rejecting immigration by suggesting that, under British Empire rule, people from British colonies had not only been encouraged but, at times forced, to live in Britain:

Did anybody force the British to buy African slaves? By buying Africans, they made them British. The same history told me that the British used extreme force to capture several African settlements. They even sent the leaders of some of these settlements into exile. An example is King JaJa of Opobo. The British traveled for Months to look for Africa. They found Africa and raped her. They have to bear the consequences of their actions. No complaints is entertained from any British about Africans (Poster P)

This then raises the issue of who has the right to call him/herself British and bestow this identity on others. The specific argument here is that African descendants have the same right to call themselves British based on their ancestral history. Here, people who reject immigration are being reminded, that, although Britain's colonial history might appear

glorious to some, British history is rooted in the (bleak) past of those living in the colonies – which, in turn, here justifies immigration.

The YouTube comments also challenged Emma's concept of 'Britishness' by referring to Britain's pre-colonial history. For example, some posters undermined Emma's notion of 'Britishness' by suggesting that:

if you look at the history of Britain [we] are pretty much all immigrants... a mix of Roman and Saxons and Norman and Viking... further back you have the Celts and the Picts... who came from somewhere else in Europe... (Poster Q)

Again, the argument here is that 'Britishness' is, in fact, a function of the nation's cultural diversity – and as this diversity is evident in ancient British history, no one should thus be considered more or less British. The implication here is that there is no inherent or fixed essence to 'Britishness'; instead, as Britain's history and present continue to be shaped by cultural diversity and multiculturalism, what it means to be British evolves and expands. This position, then, leads to an advocacy of tolerance and diversity, which is reflected in the following extract:

There are good and bad in every race. Every one has a right to their own opinion but no one has the right to disrespect and abuse people based on their race and religion. This is not freedom of speech. White people are not the only ones who contribute to this country. Everywhere you go you will see black and brown doctors, teachers, politicians, police officers and soldiers. Many have died for this country. Do you know why? Because this is our country! We are different but equal! Peace (Poster R)

Overall, as a prevalent theme, *Contesting Britishness* encapsulated attempts to define 'Britishness' – in other words, it captured efforts to define 'us'. Specifically, respondents contested the criteria that determine who should and should not qualify as British, which often related to immigration. This contestation sometimes included perceived threats to posters' sense of Britishness (i.e. social identity threat; Branscombe et al., 1999; Reicher et al., 2008). However, for many others, Britishness constituted a broad, flexible and evolving social category with immigration/multiculturalism at its very heart.

In sum, the current analysis indicates that YouTubers responded to explicit hate materials by attempting to define or contest the meaning of posters' national (i.e. here, predominantly British) identity. Responses also frequently served to position users according to their stance on racism. As such, whilst some YouTubers responded in a racist, aggressive and abusive manner, the majority of posters did not automatically endorse, or raise their responses beyond, the level of hatred displayed in the video clip.

Discussion

The objective of this study was to explore how general users of a particular SNS (i.e. YouTube) responded to online materials which could be construed as explicitly inciting hatred. To address this objective, comments in response to a YouTube video clip which depicted 'Emma', a White female train passenger racially abusing other passengers, were analysed.

Thematic Analysis (TA; Braun & Clarke, 2006) identified four superordinate themes: (1) *Making Sense of Emma*, (2) *Meeting Hatred with Hatred*, (3) *Us versus Them*, and (4) *Contesting Britishness*. *Making Sense of Emma* depicted attempts to explain Emma and her behaviour by invoking both intra-personal and social level explanations (i.e. alcohol/drug abuse, lack of intelligence, lack of educational and/or low socio-economic status) which also served to distance posters from Emma. *Meeting Hatred with Hatred* captured 'inflamed' responses to the video clip's content where hateful language

was used, including calls for symbolic punishment, to express opposition to Emma and the views embodied by her. *Us versus Them* captured posters' attempts to construct in- and out-groups in order to position themselves according to their own stance on racism. This positioning frequently occurred with reference to immigration and related to expressions of a sense of injustice and perceived relative deprivation. The final theme, *Contesting Britishness*, depicted attempts to define (or contest) the British national identity, especially in relation to Britain's colonial history and immigration. Whilst some YouTubers expressed a sense of their national identity being under threat, others suggested that there is no such thing as 'Britishness' – and, for some, immigration/ multiculturalism constituted the very essence of 'Britishness'. These themes link to existing theories as well as previous empirical findings which will be discussed below.

Posters frequently attempted to distance themselves from Emma's behaviour by positioning her as an out-group member (i.e. a racist). This, in turn, was achieved through invoking negative stereotypes (i.e. low intelligence, low socio-economic and single mother status) and the use of hateful, dehumanising language. Similar distancing strategies have been observed in previous work (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Douglas et al., 2005; Erjavec & Kovačič, 2012; Faulkner & Bliuc, 2016; Moghaddam, 2005; Mols & Jetten, 2014; Steinfeldt et al., 2010) where the use of (hateful) language has been described as a means of creating negative social comparisons and reinforcing existing stereotypes – or 'othering'. Here, this 'othering' strategy served not only as a means of *Making sense of Emma* but, paradoxically, also led to *Meeting hatred with hatred*, as expressed in Emma's dehumanisation.

The dehumanisation of perceived out-group members has been observed previously as a common linguistic 'othering' strategy (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Faulkner & Bliuc, 2016; Moghaddam, 2005). Dehumanisation of perceived out-group members has also been associated with attempts to foster a sense of moral righteousness or moral disengagement (Cammaerts, 2009; Faulkner & Bliuc, 2016) and the creation of a sense of in-group superiority (Bliuc, McGarty, Hartley, & Muntele Hendres, 2012; Mumford et al., 2008). For instance, Faulkner and Bliuc (2016) found that racists used moral disengagement strategies such as dehumanisation to justify expressions of their views. Interestingly, in this study, dehumanisation was used by both those who supported or opposed racist attitudes to justify expressing their views. In line with this, some sub-themes (i.e. *Sense of deprivation*, *Sense of injustice* and *Britishness under threat*) echo some of Faulkner and Bliuc's (2016) moral disengagement themes.

Furthermore, it has been suggested (e.g. Moghaddam, 2005) that one of the functions of dehumanisation is to enable in-group members to justify punitive treatments of individual out-group members and out-groups – which, in this study, was reflected in posters' demands for Emma's symbolic punishment. Similar calls for symbolic punishment have also been observed in other online hate contexts (e.g. Angie et al., 2011; Coffey & Woolworth, 2004; Doughty et al., 2014; Erjavec & Kovačič, 2012). Some posters' attempts to counter Emma's apparent racism with further expressions of hatred (i.e. *Meeting hatred with hatred*) seem to be in line with existing 'flaming' research (Coffey & Woolworth, 2004; Doughty et al., 2014; Lange, 2007; Moor et al., 2010). In particular, there was some evidence of 'flaming' (i.e. use of aggressive/abusive/racist language to oppose certain web content) in the online comments examined here. Specifically, posters appeared to use hateful language, or 'flames', not only to express their disagreement with the video clip, but also in order to mirror the norms conveyed in the video clip (i.e. as evidenced by Emma's own aggressive, racist, insulting and hostile language and behaviour). This corresponds to previous research on online aggression which suggests that aggressive expressions on the Internet are often a reaction to perceived aggression from others (Law, Shapka, Domene, & Gagne, 2012).

The most prominent themes identified in the current analysis – namely, *Us versus Them* and *Contesting Britishness* – explicitly tapped into social identity processes and,

as such, not only correspond to previous findings but also fit within a broader social identity approach (Blackwood et al., 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner et al., 1987; Verkuyten, 2013). For example, Angie et al. (2011) relate the use of negative stereotyping and categorical thinking – as evidenced by ‘us’ versus ‘them’ – to social identity formation. In particular, Angie et al. (2011) found the use of categorical thinking more prominent in non-violent ideological groups than in violent ones. Categorical ‘us-versus-them’ thinking was also observed by Blackwood et al. (2012), Bliuc et al. (2012) and Reicher et al. (2008). In line with previous work, in order to confirm their identification with like-minded respondents, or their group identity (i.e. here, non-racists), many posters in this study adopted similar categorical thinking strategies to those identified in previous work (e.g. Angie et al., 2011; Charteris-Black, 2006).

Posters who supported Emma’s behaviour also distinguished between ‘us’ and ‘them’ – specifically by using terminology that depicted ‘us’ as good and ‘them’ as bad. This, too, corresponds to previous observations that positive self-representation is often pitched against negative other-presentation (Douglas et al., 2005; Verkuyten, 2013) in order to create negative social comparisons which, in turn, aims to establish a sense of legitimacy and to persuade like-minded others (Charteris-Black, 2006). Here, negative other-presentations was frequently linked to expressions of a sense of injustice or relative deprivation. For example, posters who endorsed Emma’s views tended to claim that immigrants and ethnic minorities were in receipt of preferential treatment by the State or by the police – which, in turn, mirrors previous findings (e.g. Coffey & Woolworth, 2004; Douglas et al., 2005; Mols & Jetten, 2014). These posters also invoked the notion of socio-economic insecurity as a consequence of immigration; attributing blame to immigrants has previously been linked to scapegoating, disgruntled sentiments, prejudice and hate crimes towards minority groups (Bowling & Phillips, 2003; Sibbitt, 1997; Young, 1999; for a detailed review see Walters, 2011).³

Taken together, these findings closely relate to the concept of social identity threat (Branscombe et al., 1999; Reicher et al., 2008, Verkuyten, 2013) which derives from the social identity perspective (e.g. Blackwood et al., 2012, 2013, 2015; Reicher et al., 2008; Tajfel & Turner, 1979; Turner et al., 1987). Whilst there are different types of social identity threat, the current findings relate most closely to what Branscombe et al. (1999) have termed ‘distinctiveness threat’ (i.e. the fear that one’s group distinctiveness is prevented or undermined). This suggests that perceived threats to one’s group’s identity may lead to discrimination including expressions of overt racism and hostility (Pereira, Vala, & Costa-Lopes, 2010; Reicher et al., 2008; Verkuyten, 2013).

In line with previous findings (e.g. Faulkner & Bliuc, 2016), posters in this study, who supported racist views, attempted to justify their opinions by expressing a perceived sense of threat to the distinctiveness of their social (i.e. national) identity (Branscombe et al., 1999) on the basis of immigration. It has been shown previously that one of the responses to multiculturalism can be the expression of a perceived (national) identity threat (Mols & Jetten, 2014; Sindic, 2008). Issues relating to perceived social identity threat were also evident in our final theme – namely, *Contesting Britishness*. In particular, posters who supported racist views invoked as sense socio-economic threat to the White British working class.

Another way in which the themes *Us versus Them* and *Contesting Britishness* tapped into social identity processes relates to YouTubers’ contestation of their national identity as British. In particular, Faulkner and Bliuc (2016) previously observed how Australians disputed in online newspaper comments the perceived values and characteristics considered representative of their national identity. In addition, differences in people’s versions of ‘Britishness’ have been identified previously (e.g. Condor, 1997). In particular, differing accounts of ‘Britishness’ are frequently found in political discourse where they serve as rhetorical tools to influence people’s interests and behaviours

according to political agendas (Sindic, 2008). Therefore, as political agendas change, so might the boundaries/definitions of 'Britishness'.

Along similar lines, national identities are not fixed and, as Cohen (1994) has argued, their boundaries are generally 'fuzzy'. Thus, national identities are continuously re-defined through interactions with different 'others' (Condor, 2000; Hopkins & Murdoch, 1999).

Exposure to Online Hate Materials and Escalation

Coffey and Woolworth (2004) observed that some online discussions can be characterised by escalation (i.e. in terms of racism, aggression, abuse). However, in the context of the present study, the authors did not find any straightforward evidence of escalation; instead, posters' responses seemed more complex and, as discussed above, often contested relevant social categories (i.e. 'us' versus 'them', 'Britishness'). One reason for the lack of evidence of escalation here is likely to be found in the relative short time-span (i.e. 48 hours) posters had available to respond to the video clip (i.e. before the clips' removal from YouTube).

Another likely reason is the asynchronous nature of the responses under investigation here and the corresponding non-sequential way in which posts were displayed on YouTube: unlike in Coffey and Woolworth's study (2004), given the volume of simultaneous responses to the YouTube video, posters were rarely able to respond sequentially (i.e. directly) to specific others. Some respondents attempted to overcome this issue by creating a response sequence addressing other posters directly (i.e. starting their responses with an '@' and indicating the recipient's username). Yet, by the time their posts were displayed, given the sheer number and speed of postings, these attempts at creating response sequences would have been hard to identify visually (i.e. they would have appeared 'lost' amongst hundreds of other comments).

The potential differences between synchronous versus asynchronous communication and user density in online spaces have been highlighted previously (Wojcieszak, 2010). In particular, Wojcieszak (2010) questioned whether online response synchrony and user density would facilitate or prevent people from responding in a more polarised (i.e. escalated) manner over time. Thus, the relative short time, asynchrony and density of the comments analysed here might, in part, have prevented further polarisation and escalation of hateful responses. The specific relationship between response time, asynchrony, density and escalation of hateful responses therefore warrants further investigation in order to clarify the possible factors contributing to polarised responses.

Finally, whilst collating the online comments, it was observed that several respondents were removing their comments from the website. This removal seemed prompted by other respondents' appeals to the service providers (i.e. YouTube) or the police to delete or investigate certain comments for their perceived offensive and illegal nature (i.e. hate speech and/or incitement to racial hatred). Therefore, the removal of posts seemed related to posters' concerns about potential legal consequences and a decreased sense of online anonymity. More importantly, the observed removal of posts suggests that a certain degree of self-policing is practiced online.

It was also observed that some posters blocked or spammed others who expressed racist views, as well as advised others to do the same (via commenting). Consequently, responses by blocked or spammed posters became no longer visible to anyone on YouTube, suggesting that online hate materials can, in fact, elicit self-censorship. This self-censorship, in turn, seems to further support the finding that social networking users do not always automatically endorse online hate materials.

Limitations

First, as with qualitative work in general, one of the limitations of this study lies in the fact that the current findings cannot be generalised (e.g. across all online hate materials or different SNS). Here, responses to a certain, racially motivated hate incident on one specific SNS (i.e. YouTube) were explored. The authors therefore cannot claim to explain hatred across all SNSs or all hate-motivated incidents, such as those based on disability, faith, gender, transgender or subculture. Instead, the broad focus of the current research question aimed to provide a starting point for research by exploring the nature of online replies to explicit hate materials. In particular, the present study aimed to shift the currently dominant 'end-product approach' (i.e. already polarised individuals/groups and their content/interactions) to a focus on the potential processes underlying the development of polarisation in the context of online hatred. We found that whilst online racism can lead to some degree of polarisation, the majority of responses did not suggest polarisation per se, but instead indicated social identity processes at play (i.e. reflecting the construction of in- and out-groups, illustrating respondents' group affiliations).

Second, whilst one of the study's strengths lies in the realistic nature of the data, this observational approach prevents the authors from making inferences about the representativeness of the responses (e.g. the sample might not represent users across other SNSs). The authors neither had access to respondents' demographics, nor could they determine whether respondents posted with multiple or different user names (i.e. identities). Yet, given the substantial number of comments in response to the video clip (i.e. over 71,000 posts), it is unlikely that these were exclusive to a specific (demographic) group – instead, there was evidence of some diversity in backgrounds (e.g. gender, ethnicity, geographic location).

Finally, as in previous studies (e.g. Coffey & Woolworth, 2004), the findings do not allow the authors to draw inferences about the effects that the relatively short period of time in which people were able to respond to the video clip might have had on the content of their posts. In particular, it remains unclear whether longer-term online interactions result in repeated and/or increased expressions of hatred or whether long-term interactions/discussions change direction and automatically lose focus, therefore reducing or diverting expressions of hatred.

Conclusion

Expressions of hatred (e.g. racism) are more and more becoming a feature of life online, especially on social networking sites (e.g. YouTube). It is therefore becoming increasingly important to understand how 'everyday' online users respond to and engage with such hateful materials. Many YouTubers who responded to the hateful online material analysed here condemned its content. As such, the findings obtained in the current context indicate that online hatred may not always be automatically endorsed by online users; instead, we observed a degree of contestations in online users' responses. Whilst the current findings cannot be generalised, we nevertheless hope they provide a springboard for future research which should take into account not only a wider variety of hate materials but also different social networking sites.

Notes

1. This comment was made in response to another comment, expressing opposition to the poster's views.

2. Please note that this comment was also made in response to another comment. In this instance, the comment was aimed at a poster who shared the views of this particular poster.

3. At the time of the video clip's upload (e.g. October 2011), Britain was in an economic recession and characterised by socio-economic insecurity. This may therefore have exacerbated posters' sense of relative material and financial deprivation.

References

- AFP. (2016, March 8). Australian behind Singapore hate site convicted of sedition. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/wires/afp/article-3482015/Australian-Singapore-hate-site-convicted-sedition.html>
- Agarib, A. (2016, January 1). Man arrested for online hate crime. *Khaleej Times*. Retrieved from <http://www.khaleejtimes.com/article/20160101/ARTICLE/301019214>
- Allen, M., Bromley, A., Kuyken, W., & Sonnenberg, S. (2009). Participants' experiences of mindfulness-based cognitive therapy: "It changed me in about every way possible". *Behavioural and Cognitive Psychotherapy*, *37*, 413-430. <http://dx.doi.org/10.1017/S135246580999004X>
- Angie, A. D., Davis, J. L., Allen, M. T., Byrne, C. L., Ruark, G. A., Cunningham, C. B., . . . Mumford, M. D. (2011). Studying ideological groups online: Identification and assessment of risk factors for violence. *Journal of Applied Social Psychology*, *41*, 627-657. <http://dx.doi.org/10.1111/j.1559-1816.2011.00730.x>
- Anti-Defamation League. (2005, May). *Attitudes toward Jews in twelve European countries*. Retrieved from http://www.adl.org/israel/Eur_Poll_Israel_May_2005.pdf
- Baron, N. S. (2004). See you online: Gender issues in college students use of instant messaging. *Journal of Language and Social Psychology*, *23*, 397-423. <http://dx.doi.org/10.1177/0261927X04269585>
- BBC News. (2012, February 10). Derby men jailed for giving out gay death call leaflets. *BBC News*. Retrieved from <http://www.bbc.co.uk/news/uk-england-derbyshire-16985147>
- BBC News. (2016, January 12). Teenager given probation for racist slogan. *BBC News*. Retrieved from <http://www.bbc.com/news/uk-northern-ireland-35294697>
- Blackwood, L., Hopkins, N., & Reicher, S. (2012). Divided by a common language? Conceptualising identity, discrimination and alienation. In K. J. Jonas & T. A. Morton (Eds.), *Restoring civil societies: The psychology of intervention and engagement following crisis* (pp. 222-236). Chichester: John Wiley & Sons Inc.
- Blackwood, L., Hopkins, N., & Reicher, S. (2013). I know who I am, but who do they think I am? Muslim perspectives on encounters with airport authorities. *Ethnic and Racial Studies*, *36*, 1090-1108. <http://dx.doi.org/10.1080/01419870.2011.645845>
- Blackwood, L., Hopkins, N., & Reicher, S. (2015). Turning the gaze on 'us': The role of authorities in the alienation of minorities. *European Psychologist*, *18*, 245-252. <http://dx.doi.org/10.1027/1016-9040/a000151>
- Blazak, R. (2001). White boys to terrorist men: Target recruitment of Nazi skinheads. *American Behavioral Scientist*, *44*, 982-1000. <http://dx.doi.org/10.1177/00027640121956629>

- Bliuc, A. M., McGarty, C., Hartley, L. K., & Muntele Hendres, D. (2012). Manipulating national identity: the strategic use of rhetoric by supporters and opponents of the 'Cronulla riots' in Australia. *Ethnic and Racial Studies*, *35*, 2174-2194. <http://dx.doi.org/10.1080/01419870.2011.600768>
- Bostdorff, D. M. (2004). The internet rhetoric of the Ku Klux Klan: A case study in website community building run amok. *Communication Studies*, *55*, 340-361. <http://dx.doi.org/10.1080/10510970409388623>
- Bowling, B., & Phillips, C. (2003). Racist victimization in England and Wales. In D. Hawkins (Ed.), *Violent crime, assessing race & ethnic difference* (pp. 154-170). Cambridge: Cambridge University Press.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The content and context of social identity threat. In R. Spears, B. Doosje, & N. Ellemers (Eds.), *Social identity: Context, commitment, content* (pp. 35-55). Oxford: Blackwell.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101. <http://dx.doi.org/10.1191/1478088706qp063oa>
- Burrows, T. (2016, March 23). Talent agency boss who 'confronted a Muslim woman in London and asked her to explain Brussels' is arrested for 'inciting racial hatred'. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/news/article-3506275/publicity-boss-asked-Muslim-woman-street-explain-Brussels-deluged-angry-hilarious-Tweets-reply.html>
- Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces: The case of north Belgian extreme right discourses. *International Journal of Cultural Studies*, *12*, 555-575. <http://dx.doi.org/10.1177/1367877909342479>
- Castle, C. (2013, October 27). YouTube awards tarnished by YouTube hate videos and jihadi recruitment. *Music, Technology, Policy*. Retrieved from <https://musictechpolicy.wordpress.com/2013/10/27/youtube-awards-tarnished-by-youtube-hate-videos-and-jihadi-recruitment/>
- Charteris-Black, J. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society*, *17*, 563-581. <http://dx.doi.org/10.1177/0957926506066345>
- Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies*, *65*, 57-70. <http://dx.doi.org/10.1016/j.ijhcs.2006.08.009>
- Citron, D. K., & Norton, H. L. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review*, *91*, 1435-1484, University of Maryland Legal Studies Research Paper No. 2011-16.
- Coffey, B., & Woolworth, S. (2004). "Destroy the scum, and then neuter their families:" The web forum as a vehicle for community discourse? *The Social Science Journal*, *41*, 1-14. <http://dx.doi.org/10.1016/j.soscij.2003.10.001>
- Cohen, R. (1994). *Frontiers of Identity: The British and others*. London: Longman.
-

- Condor, S. (1997). Having history: A social psychological exploration of Anglo-British stereotypes. In C. C. Barfoot (Ed.), *Beyond pug's tour: National and ethnic stereotyping in theory and literature practice* (pp. 213-253). Amsterdam: Rodopi.
- Condor, S. (2000). 'Pride and prejudice': Identity management in English people's talk about "this country". *Discourse & Society, 11*, 175-205. <http://dx.doi.org/10.1177/0957926500011002003>
- Crown Prosecution Service. (2010, February 17). Teenager admits incitement to racial hatred over YouTube videos. *CPS.ORG.UK*. Retrieved from http://www.cps.gov.uk/news/latest_news/teenager_admits_incitement_to_racial_hatred/
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. <http://dx.doi.org/10.1037/h0047358>
- De Koster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me'. *Information, Communication & Society, 11*, 1155-1176. <http://dx.doi.org/10.1080/13691180802266665>
- Doane, A. (2006). What is racism? Racial discourse and racial politics. *Critical Sociology, 32*, 255-274. <http://dx.doi.org/10.1163/156916306777835303>
- Doughty, M., Lawson, S., Linehan, C., Rowland, D., & Bennett, L. (2014). Disinhibited abuse of othered communities by second-screening audiences. *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video* (pp. 55-62). ACM. <http://dx.doi.org/10.1145/2602299.2602311>
- Douglas, K. M. (2007). Psychology, discrimination and hate groups online. In A. Joinson, K. McKenna, T. Postmes, & U. D. Reips (Eds.), *The Oxford handbook of internet psychology* (pp. 155-163). New York: Oxford University Press.
- Douglas, K. M., McGarty, C., Bliuc, A. M., & Lala, G. (2005). Understanding cyberhate: Social competition and group creativity in online white supremacy groups. *Social Science Computer Review, 23*, 68-76. <http://dx.doi.org/10.1177/0894439304271538>
- Erjavec, K., & Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society, 15*, 899-920. <http://dx.doi.org/10.1080/15205436.2011.619679>
- ETHNOS. (2005). *Citizenship and belonging: What is Britishness?* Commission for Racial Equality (CRE). London: Ethnos Research and Consultancy.
- Ezekiel, R. S. (2002). An ethnographer looks are neo-Nazi and Klan groups: The racist mind revisited. *American Behavioral Scientist, 46*, 51-71. <http://dx.doi.org/10.1177/0002764202046001005>
- Faulkner, N., & Bliuc, A. M. (2016). "It's okay to be racist": Moral disengagement in online discussions of racist incidents in Australia. *Ethnic and Racial Studies, 39*, 2545-2563. <http://dx.doi.org/10.1080/01419870.2016.1171370>
- Foxman, A. H., & Wolf, C. (2013). *Viral hate containing its spread on the internet*. New York: Palgrave Macmillan.

- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400. [http://dx.doi.org/10.1016/0191-8869\(86\)90014-0](http://dx.doi.org/10.1016/0191-8869(86)90014-0)
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193. <http://dx.doi.org/10.1111/j.1745-6916.2007.00036.x>
- Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, 3, 29-44. <http://dx.doi.org/10.1111/j.1530-2415.2003.00013.x>
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58, 177-193. <http://dx.doi.org/10.1111/1540-4560.00255>
- Gleibs, I. H., Sonnenberg, S., & Haslam, C. (2014). We get to decide": The role of collective engagement in counteracting feelings of confinement and lack of autonomy in residential care. *Activities, Adaptation & Aging*, 38, 259-280. <http://dx.doi.org/10.1080/01924788.2014.966542>
- Green, D. P., Abelson, R. P., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice & D. T. Miller (Eds.), *Cultural Divides* (pp. 429-464). New York: Russell Sage Foundation.
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6, 215-242. <http://dx.doi.org/10.1515/jplr.2010.011>
- Herring, S., Job-Sluder, K., Scheckler, R., & Brab, S. (2002). Searching for safety online: "Trolling" in a feminist forum. *The Information Society: An International Journal*, 18, 371-384. <http://dx.doi.org/10.1080/01972240290108186>
- Hopkins, N., & Murdoch, N. (1999) The role of the 'other' in national identity: Exploring the context-dependency of the national in-group stereotype. *Journal of Community and Applied Social Psychology*, 9, 321-338. [http://dx.doi.org/10.1002/\(SICI\)1099-1298\(199909/10\)9:5<321::AID-CASP510>3.0.CO;2-M](http://dx.doi.org/10.1002/(SICI)1099-1298(199909/10)9:5<321::AID-CASP510>3.0.CO;2-M)
- Internet Live Stats. (2016a, 21 March). *Internet users*. Retrieved from <http://www.internetlivestats.com/internet-users/>
- Internet Live Stats. (2016b, 21 March). *Facebook users*. Retrieved from <http://www.internetlivestats.com/>
- Jacks, W., & Adler, J. R. (2015). A proposed typology of online hate crime. *Open Access Journal of Forensic Psychology*, 7, 64-89. Retrieved from <http://eprints.mdx.ac.uk/18930/1/Jacks%20and%20Adler%202015.pdf>
- Joinson, A. N. (1998). Causes and implication of disinhibited behaviour on the internet. In J. Gackenbach (Ed.), *Psychology and the internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43-60). San Diego: Academic Press.
- Kazakhstan Newsline. (2015, August 3). New person sentenced for 'inciting religious hatred' in Kazakhstan. *Kazakhstan Newsline*. Retrieved from <http://www.newsline.kz/news.php?y=15&m=08&d=03#0>

- Lange, P. G. (2007). *Commenting on comments: Investigating responses to antagonism on YouTube*. Paper presented at the Annual Conference of the Society for Applied Anthropology, Tampa, Florida. Retrieved from <https://www.yumpu.com/en/document/view/14888058/commenting-on-comments-investigating-responses-to-antagonism->
- Latest.com. (2015, January 12). *Kazakhstani woman facing prison time after voicing her pro-Russian stance*. Retrieved from <http://latest.com/2015/01.kazakhstani-woman-facing-prison-time-voicing-pro-russian-stance/>
- Law, D. M., Shapka, J. D., Domene, J. F., & Gagne, M. H. (2012). Are cyberbullies really bullies? An investigation of reactive and proactive online aggression. *Computer in Human Behavior, 28*, 664-672. <http://dx.doi.org/10.1016/j.chb.2011.11.013>
- Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist, 45*, 927-957. <http://dx.doi.org/10.1177/0002764202045006003>
- Levin, B. (2002). Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioral Scientist, 45*, 958-988. <http://dx.doi.org/10.1177/0002764202045006004>
- McNab, H. (2015, February 12). Crucifixions and a \$50 living allowance: Aussie who joined ISIS reveals what life is really like in Syria after defecting because he didn't like Westerners being executed. *Daily Mail Australia*. Retrieved from <http://www.dailymail.co.uk/news/article-2950469/It-s-not-like-YouTube-Former-ISIS-member-defected-terror-group-reveals-escaped-brutal-executions-Westerners-forced-out.html>
- McNamee, L. G., Peterson, B. L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs, 77*, 257-280. <http://dx.doi.org/10.1080/03637751003758227>
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist, 60*, 161-169. <http://dx.doi.org/10.1037/0003-066X.60.2.161>
- Mols, F., & Jetten, J. (2014). No guts, no glory: How framing the collective past paves the way for anti-immigrant sentiments. *International Journal of Intercultural Relations, 43*, 74-86. <http://dx.doi.org/10.1016/j.ijintrel.2014.08.014>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior, 26*, 1536-1546. <http://dx.doi.org/10.1016/j.chb.2010.05.023>
- Mumford, M. D., Bedell-Avers, K. E., Hunter, S. T., Espejo, J., Eubanks, D., & Connelly, M. S. (2008). Violence in ideological and non-ideological groups: A quantitative analysis of qualitative data. *Journal of Applied Social Psychology, 38*, 1521-1561. <http://dx.doi.org/10.1111/j.1559-1816.2008.00358.x>
- Nirvana News. (2016, January 5). Trump used in terrorist recruitment video. *Nirvana News*. Retrieved from <https://www.youtube.com/watch?v=cv1aU3pG3BQ>
-

- Osborn, A., & Dyomkin, D. (2015, December 30). Russian court jails blogger for five years for 'extremist' posts. *Reuters*. Retrieved from <http://uk.mobile.reuters.com/article/idUKKBN0UD16O20151230>
- O'Sullivan, P. B., & Flanagan, A. (2003). Reconceptualizing "flaming" and other problematic communication. *New Media and Society* 5, 69-94. <http://dx.doi.org/10.1177/1461444803005001908>
- Parisi, P. (1998). A sort of compassion: The Washington Post explains the crisis in urban America. *Howard Journal of Communications*, 9, 187-203. <http://dx.doi.org/10.1080/106461798246989>
- Peddinti, S., Ross, K., & Cappos, J. (2014). On the internet, nobody knows you're a dog: A Twitter case study of anonymity in social networks. In *Proceedings of the second ACM conference on Online social networks - COSN '14* (pp. 83-94). ACM. <http://dx.doi.org/10.1145/2660460.2660467>
- Pereira, C., Vala, J., & Costa-Lopes, R. (2010). From prejudice to discrimination: The legitimizing role of perceived threat in discrimination against immigrants. *European Journal of Social Psychology*, 40, 1231-1250. <http://dx.doi.org/10.1002/ejsp.718>
- Perry, B. (2000). "Button-down terror": The metamorphosis of the hate movement. *Sociological Focus*, 33, 113-131. <http://dx.doi.org/10.1080/00380237.2000.10571161>
- Phillips, T. (2015, December 22). Pu Zhiqiang given three-year suspended sentence. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2015/dec/22/pu-zhiqiang-chinese-human-rights-lawyer-sentenced-to-three-years>
- Poch, R. (2015, November 15). Fighting incitement on social media, "they can't" gets 40,000 hate videos removed from YouTube. *Breaking Israel News*. Retrieved from <http://www.breakingisraelnews.com/53013/they-cant-launches-campaign-against-palestinian-incitement-social-media-terror-watch/#toBExsVzYqfs3VLg.97>
- Radio Free Europe Radio Liberty. (2015, July 31). Another Kazakh sentenced for inciting ethnic hatred online. *Radio Free Europe Radio Liberty*. Retrieved from <http://www.rferl.org/a/kazakhstan-ethnic-hatred-man-sentenced/27163239.html>
- Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior*, 29, 610-615. <http://dx.doi.org/10.1016/j.chb.2012.11.011>
- Reicher, S., Haslam, S. A., & Rath, R. (2008). Making a virtue of evil: A five-step social identity model of the development of collective hate. *Social and Personality Psychology Compass*, 2, 1313-1344. <http://dx.doi.org/10.1111/j.1751-9004.2008.00113.x>
- Risen, T. (2014, September 5). Twitter, Facebook, YouTube grapple with Islamic State censorship. *US News*. Retrieved from <http://www.usnews.com/news/articles/2014/09/05/twitter-facebook-youtube-navigate-islamic-state-censorship>
-

- Robi, E. (2013). What computer forensics tell us about the Boston bombers. *Elluma Discovery*. Retrieved from <http://www.elluma.com/what-computer-forensics-tells-us-about-the-boston-bombers/>
- Runciman, W. G. (1966). *Relative deprivation and social justice*. London: Routledge.
- Sibbitt, R. (1997). *The perpetrators of racial harassment and racial violence*. London: Home Office.
- Sindic, D. (2008). National identities: Are they declining? *Beyond Current Horizons: Technology, Children, Schools and Families*, 1-22. Retrieved from http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/final_sindic_nationalidentities_20081201.pdf
- Spencer, R. (2015, December 31). Kazakhstan: Convert from Islam to Christianity gets 2 years prison for inciting religious hatred. *Jihad Watch*. Retrieved from <https://www.jihadwatch.org/2015/12/kazakhstan-convert-from-islam-to-christianity-gets-2-years-prison-for-inciting-religious-hatred>
- Steinfeldt, J. A., Foltz, B. D., Kaladow, J. K., Carlson, T. N., Pagano, L. A., Jr., Benton, E., & Steinfeldt, M. C. (2010). Racism in the electronic age: Role of online forums in expressing racial attitudes about American Indians. *Cultural Diversity and Ethnic Minority Psychology*, 16, 362-371. <http://dx.doi.org/10.1037/a0018692>
- Stern, K. S. (2002). *Hate and the internet*. New York: The American Jewish Committee.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7, 321-326. <http://dx.doi.org/10.1089/1094931041291295>
- Swinford, S. (2013, May 26). Woolwich attack: Google must block hate preachers, says Theresa May. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/10081617/Woolwich-attack-Google-must-block-hate-preachers-says-Theresa-May.html>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Pacific Grove, CA: Books/Cole.
- Thomas, S. (1998). Race, gender, and welfare reform: The antinationalist response. *Journal of Black Studies*, 28, 419-446.
- Turnage, A. K. (2008). Email flaming behaviors and organizational conflict. *Journal of Computer-Mediated Communication*, 13, 43-59. <http://dx.doi.org/10.1111/j.1083-6101.2007.00385.x>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A Self-Categorization Theory*. Oxford & New York: Blackwell.
- Turpin-Petrosino, C. (2002). Hateful sirens . . . who hears their song? An examination of student attitudes toward hate groups and affiliation potential. *Journal of Social Issues*, 58, 281-301. <http://dx.doi.org/10.1111/1540-4560.00261>
-

Verkuyten, M. (2013). Justifying discrimination against Muslim immigrants: Out-group ideology and the five-step social identity model. *British Journal of Social Psychology*, 52, 345-360. <http://dx.doi.org/10.1111/j.2044-8309.2011.02081.x>

Walters, M. A. (2011). A general theories of hate crime? Strain, doing difference and self control. *Critical Criminology* 19, 313-330. <http://dx.doi.org/10.1007/s10612-010-9128-2>

Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogenous online groups and politically dissimilar offline ties on extremism. *New Media & Society* 12, 637-655. <http://dx.doi.org/10.1177/1461444809342775>

Young, J. (1999). *The exclusive society: Social exclusion, crime and difference in late modernity*. London: Sage Publications.

YouTube (n.d.). *YouTube-statistics*. Retrieved from <https://www.youtube.com/yt/press/en-GB/statistics.html>

Zhou, Y., Reid, E., Qin, J., H. Chen, H., & Lai, G. (2005). US domestic extremist groups on the web: Link and content analysis. *IEEE Intelligent Systems*, 20, 44-51. <http://dx.doi.org/10.1109/MIS.2005.96>
